

Accurate Fitting of Noisy Irregular Beam Data for the Planck Space Telescope

Oscar Borries*, Per Heighwood Nielsen*, Jan Tauber[†], Arturo Martín-Polegre[†], Hans Bruun Nielsen[‡]
*TICRA

Læderstræde 34, 1201 Copenhagen, Denmark
E-mail: ob@ticra.com, phn@ticra.com (corresponding author)
[†]ESTEC

Keplerlaan 1, 2200 AG Noordwijk, The Netherlands
E-mail: Jan.Tauber@esa.int, Arturo.Martin.Polegre@esa.int
[‡]DTU Informatics, Technical University of Denmark
Richard Petersens Plads, Building 321, 2800 Lyngby, Denmark
E-mail: hbn@imm.dtu.dk

Abstract—Accurate fitting of the noisy irregular amplitude only main beam data is essential for the retrieval of the Planck space telescope geometry [1]. Therefore, a novel two-step fitting algorithm which focuses on the spatial dependency of the in-flight measurements has been implemented. To reduce both the noise and the size of the dataset a spatial filter is applied, without reducing the amount of pattern information. Thereafter, a Kriging [2], [3] fitting is performed, providing a smooth model with a significant noise level reduction. As a result, this algorithm provides a much more accurate and smoother result, reasonable error estimates and runtimes several orders of magnitudes faster than the previous algorithms.

Index Terms—Fitting, Planck, Geometry Retrieval, Kriging, Beam Data.

I. INTRODUCTION

The in-flight measurements from the Planck satellite are impacted by severe noise, particularly for the Low-Frequency Instruments (LFI). Furthermore, they are obtained in an irregular pattern as a consequence of the rotation performed by the satellite and signals measured several times in nearly the same directions.

An example of a dataset¹ covering the main beam from the 30 GHz LFI detector is shown in Figure 1, illustrating the distribution of the 41677 samples in "lines" along the scan direction.

¹Note that all data used in this paper are simulated.

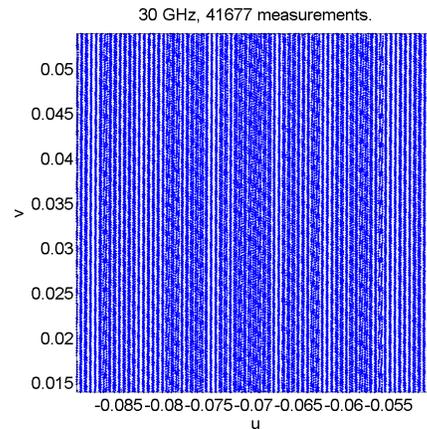


Fig. 1. An example of a dataset from main beam of the Planck detectors. Note the distribution in lines along the scan direction.

To reduce the noise, and achieve a model in a regular grid, a two-stage fitting algorithm has been developed, exploiting the spatial dependency of the measurements. The first stage consists of a rather crude filter, the purpose of which is both to reduce the noise and the size of the dataset. The second stage is the Kriging fitting model, fitting the result of the filter to provide a smooth model. This stage is inspired by the implementation in the MATLAB Kriging toolbox DACE [3], [4], modified to employ fitting and implemented in FORTRAN with a focus on memory-efficiency and stability.

II. ALGORITHM

Mathematically, we are given \bar{m} measurements of (u, v, z) , with the (u, v) coordinates of each measurement are called the *sites*, and z are called the *responses*. As mentioned previously, the samples are noisy and irregularly distributed, but have a spatial dependency, such that the closer two measurements are in the (u, v) plane, the greater the correlation between their responses.

With the datasets we are considering, \bar{m} is far too large to employ the Kriging algorithm directly - as will be demonstrated in Section II-B1, the chief computational cost of computing the Kriging model is the Cholesky factorization of a matrix of size $m \times m$, where m is the number of sites. For the 30 GHz dataset, $\bar{m} = 41677$, and thus some data reduction is needed. Furthermore, experiments have shown that Kriging fitting perform poorly on datasets affected by serious noise, particularly for low-dimensional data [5, Chapter 4].

A. Filter

Therefore, a crude spatial filter is employed. The filter separates the (u, v) space into a grid, and uses the mean of all measurements in each grid-block - an illustration is shown in Figure 2, and further discussion in [6], [7]. Thus, assuming that the noise has mean and skewness zero, taking the mean of the responses z in each grid-block should allow us to filter out most of the noise. Of course, we need a sufficient number of measurements in each block, thus encouraging us to make the blocks as large as possible - on the other hand, since we only get one resulting measurement per block to input to our Kriging model, too large blocks will result in too few measurements and consequently a result that does not take into account the local variations (such as the shoulders of the main beam).

To investigate this, we considered the spacing between the regular grid as a multiplier κ times the Nyquist ratio λ/D , λ being the wavelength and D being the aperture of the antenna. Previous experience with representing fields by interpolation suggested using $\kappa \approx 0.25$, but as discussed in detail in [7], the noisy data in the problem resulted

in $0.05 \leq \kappa \leq 0.1$ yielding far better solutions, depending on the dataset in question².

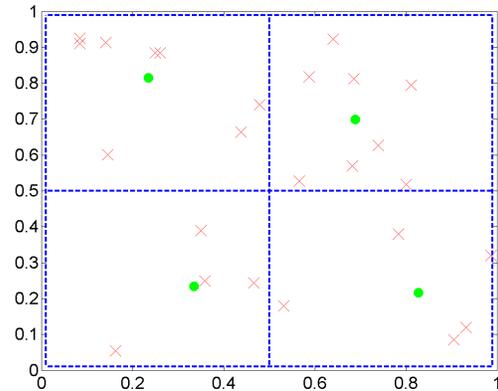


Fig. 2. A mini-example of the filter. The blue lines represent the grid, separated by $\kappa \frac{\lambda}{D}$. The red crosses are the measurements, and the green points are the result of the filter - representing the mean of the samples inside each box.

B. Kriging

The method of Kriging exploits a supposed spatial dependency in a sample to impose additional requirements on the fit. In its simplest form, it basically only involves the fitting of a correlation model to a sample set - it was in this form the Danie Krige [2] introduced it. Later work by several people, most notably G. Matheron [8], formalized it further and introduced several variations of the model, including the *Universal Kriging* model used here.

Its use in modelling deterministic behaviour was introduced by the landmark paper [9], allowing widespread use of the method which was previously restricted to the geostatistical community. Theoretically, the key strength of the Kriging predictor is that amongst all linear and unbiased estimators, it minimizes the expected error³. In practice, it has several other advantages which has prompted its use in the present scenario - most notably, it yields a smooth model and requires no special considerations when faced with irregular data. Also, its use of a global regression model and

²Specifically, this depends on the detector in question - the lower the frequency, the more noisy the dataset.

³Technically, it is a BLUP: Best Linear Unbiased Predictor, where "Best" refers to it minimizing the mean squared error [10, p. 60].

a local correlation model allows for surprisingly good accuracy when used correctly.

1) *Model*: Given a set of m design sites $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^T$, $\mathbf{x}_i \in \mathbb{R}^2$, $\mathbf{x}_i = (u_i, v_i)$ and responses $\mathbf{z} = [z_1, z_2, \dots, z_m]^T$, we begin by *normalizing* the data by subtracting the mean and dividing by the standard deviation, for each column of X and for \mathbf{z} . In this manner, each column of X and \mathbf{z} have a mean of zero and a standard deviation of one [3, eqn. (2.1)]. This process yields better numerical and statistical properties [10, p. 139].

Inspired by [9], a model is adopted that expresses the deterministic response $\mathbf{z}(\mathbf{x})$ as the realization of a regression model \mathcal{F} and a stochastic model \mathcal{S} , such that our model of the responses \mathbf{z} is

$$\mathbf{z}(\mathbf{x}) = \mathcal{F}(\mathbf{x}, \boldsymbol{\beta}) + \mathcal{S}(\mathbf{x}, \boldsymbol{\theta}) \quad (\text{II.1})$$

In the present scenario, the regression model is restricted to a linear⁴ combination of n basis functions f_j , specifically low-order polynomials. $\boldsymbol{\beta} \in \mathbb{R}^n$ are the *regression parameters*, acting as weights for the basis functions. n depends on the order of the polynomials chosen - in our experiments, we found that the use of second order polynomials⁵ provided the best result, yielding $n = 6$. The stochastic model is a Gaussian correlation model (II.4), and $\boldsymbol{\theta} \in \mathbb{R}^2$ act as *scaling parameters*, modifying the amount of correlation relative to the distance between two sites. This allows us to express (II.1) as

$$\mathbf{z} = F\boldsymbol{\beta} + \Phi(\boldsymbol{\theta}, \gamma)\boldsymbol{\alpha} \quad (\text{II.2})$$

where F is the *regression matrix*, i.e. the basis functions evaluated at the given sites $F_{ij} = f_j(\mathbf{x}_i)$, $i = [1, 2, \dots, m]$, $j = [1, 2, \dots, n]$. And the *correlation matrix* Φ is

$$\Phi_{ij} = \mathcal{Z}(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_j), \quad i, j = 1, 2, \dots, m \quad (\text{II.3})$$

As mentioned, we use a Gaussian correlation model $r(\theta_k, x_{i,k}, x_{j,k})$, such that

$$r(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_k) = \prod_{k=1}^2 e^{-\theta_k(x_{i,k} - x_{j,k})^2} \quad (\text{II.4})$$

meaning that $\Phi_{ij} = r(\boldsymbol{\theta}, \mathbf{x}_i, \mathbf{x}_k)$. Here, the notation $x_{i,k}$ means the k 'th coordinate of the i 'th

⁴In theory, there is no need to restrict the regression model to being linear - however, a non-linear model would require a substantial amount of work to implement, and would increase the computation time significantly.

⁵Implemented as the `regpoly2` regression model in DACE, [3, Section 5.3].

site. To achieve a fitting model, we add a constant $\gamma > 0$ to the diagonal of Φ , yielding a final model

$$\mathbf{z} = F\boldsymbol{\beta} + (\Phi(\boldsymbol{\theta}) + \gamma I)\boldsymbol{\alpha} \quad (\text{II.5})$$

Note that the i 'th diagonal element of Φ is the autocorrelation of the i 'th site - for interpolating Kriging, this would be 1, but when fitting, we set the autocorrelation to $1+\gamma$, giving the model more freedom to follow a more likely path.

Computing the parameters, namely $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and γ , is by far the most tricky aspect of the implementation, as great care needs to be taken to ensure numerically stable and computationally efficient results. This is detailed in several papers, amongst them [3], [4], [6] and in the book [11], and will therefore be skipped here. Thus, the key thing to note is how the model (II.5) consists of a global regression part and a local correlation part.

2) *Predictor*: Having computed the model, we can predict the value $\hat{z}(\hat{\mathbf{x}})$ at an untried location $\hat{\mathbf{x}}$ as [3, (2.16)]

$$\hat{z}(\hat{\mathbf{x}}) = \mathbf{f}(\hat{\mathbf{x}})^T \boldsymbol{\beta} + \mathbf{r}(\boldsymbol{\theta}, \hat{\mathbf{x}}, X)^T \boldsymbol{\alpha} \quad (\text{II.6})$$

where $\mathbf{r}(\boldsymbol{\theta}, \hat{\mathbf{x}}, X) \in \mathbb{R}^m$ is the value of the correlation function (II.4) evaluated between each of the m given sites and the untried $\hat{\mathbf{x}}$. This yields a prediction in the normalized space - to scale back to the original space, i.e. before the transformation mentioned at the beginning of Section II-B1, simply multiply by the standard deviation of the measured response vector \mathbf{z} .

III. RESULTS

The 30 GHz beam is simulated with a noise level of 14 dB below peak, which is generated by the sampled planet and the temperature of the detector [1]. The main beam pattern is calculated by Physical Optics [12] in all the 41677 measurement directions and the resulting simulated pattern is shown in the 3D view in Figure 3.

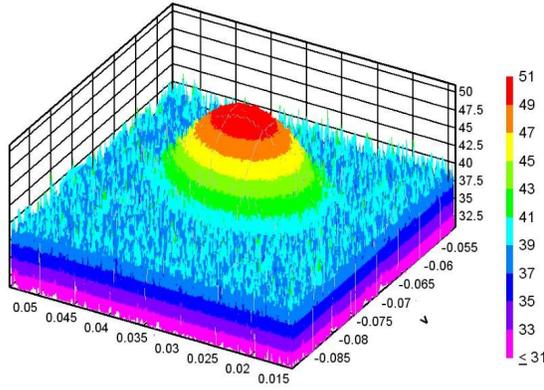


Fig. 3. Simulated main beam with noise for 30 GHz detector LFI27S.

Applying the algorithm with $\kappa = 0.1$, and Kriging parameters as computed by the algorithm (note that θ and γ are computed automatically by a Maximum Likelihood Estimate [4, Section 6]) yields the result shown in Figure 4.

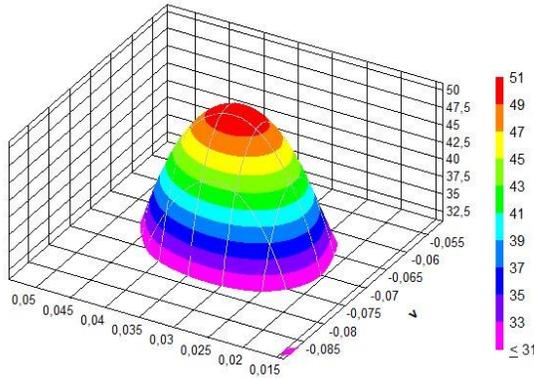


Fig. 4. Kriging fitted and noise filtered main beam for 30 GHz detector LFI27S.

The fitted simulated main beam is compared with the noise-less simulated beam in Figure 5. The main shapes of the contour curves down to the introduced noise level of 14 dB are very well regenerated. Below -25 dB the noise is dominant and all pattern information is lost in this region.

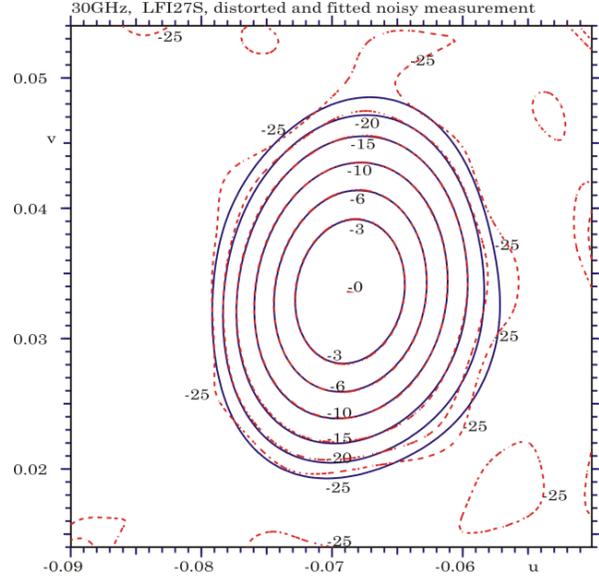


Fig. 5. Kriging fitted and noise filtered main beam, shown in dashed red curves, compared with noiseless pattern for 30 GHz detector, LFI27S, shown in solid blue curves.

The Kriging fitted beams are used directly in the retrieval of the geometrical information on the Planck telescope. Due to the very exact fitting and noise reduction a larger dynamic range of the measured beams can be utilized in the retrieval giving much more information on the telescope mirror alignment and surface deformations. Furthermore, a Kriging generated regular grid makes it much simpler to calculate the main beam characteristics, such as beam peak, half power ellipticity and efficiency.

IV. CONCLUSION

The algorithm developed for this project makes use of a crude filter and a fitting Kriging model (II.5), exploiting the spatial dependency of the data. The results are quite impressive - the severe noise level is significantly reduced, allowing a larger range of the measured beams to be utilized in the geometry retrieval. Furthermore, although not discussed here, the runtimes are very reasonable and on the order of seconds, allowing a more interactive approach to the modelling process. In conclusion, the algorithm presented here improves upon previous algorithms and will be vital in the in-flight geometry retrieval of the Planck space telescope.

REFERENCES

- [1] F. Jensen, P. H. Nielsen, J. Tauber, and A. Martín-Polegre, “In-Flight Retrieval of Reflector Anomalies for the Planck Space Telescope,” in *EUCAP*, 2010.
- [2] D. G. Krige, “A statistical approach to some basic mine valuation problems on the Witwatersrand,” *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, pp. 119–139, 1951.
- [3] Søren N. Lophaven, H. Nielsen, and Jacob Søndergaard, “DACE, a MATLAB Kriging toolbox, Version 2.0,” IMM, DTU, Tech. Rep. IMM-REP-2002-12, 2002, available online at <http://www2.imm.dtu.dk/~hbn/dace/dace.pdf>.
- [4] —, “Aspects of the MATLAB toolbox DACE,” IMM, DTU, Tech. Rep. IMM-REP-2002-13, 2002, available online at <http://www.imm.dtu.dk/~hbn/publ/TR0213.ps>.
- [5] Oscar Borries, “Surrogate Modelling using DACE,” 2009, Available from the Orbit database, <http://tiny.cc/2xeaa>.
- [6] Oscar Borries and Per Heighwood Nielsen, “In-flight Retrieval of Geometrical Information on the Planck Telescope,” TICRA, Tech. Rep. S-1531-08, 2010.
- [7] Oscar Borries, “Numerical Approximation of Surfaces,” Technical University of Denmark, Tech. Rep., 2010.
- [8] G. Matheron, “Principles of Geostatistics,” *Economic Geology*, pp. 1246–1266, 1963.
- [9] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, “Design and analysis of computer experiments,” *Stat. Sci.*, vol. 4, pp. 409–435, 1989.
- [10] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*, 1st ed., ser. Springer Series in Statistics. Springer, 2003.
- [11] Alexander Forrester, András Sobester, and Andy Keane, *Engineering Design via Surrogate Modelling*, 1st ed. Wiley, 2007.
- [12] Knud Pontoppidan, *GRASP9 Technical Description*, 1st ed. TICRA, 2005.