

Astrophysical data analysis with information field theory

Torsten Enßlin

*Max Planck Institut für Astrophysik (Karl-Schwarzschild-Straße 1, D-85748 Garching, Germany), and
Ludwig-Maximilians-Universität München (Geschwister-Scholl-Platz 1, D-80539 München, Germany)*

Abstract. Non-parametric imaging and data analysis in astrophysics and cosmology can be addressed by information field theory (IFT), a means of Bayesian, data based inference on spatially distributed signal fields. IFT is a statistical field theory, which permits the construction of optimal signal recovery algorithms. It exploits spatial correlations of the signal fields even for nonlinear and non-Gaussian signal inference problems. The alleviation of a perception threshold for recovering signals of unknown correlation structure by using IFT will be discussed in particular as well as a novel improvement on instrumental self-calibration schemes. IFT can be applied to many areas. Here, applications in cosmology (cosmic microwave background, large-scale structure) and astrophysics (galactic magnetism, radio interferometry) are presented.

INFORMATION FIELD THEORY

Information field theory (IFT) is information theory for fields, describing in mathematical language how information on spatially distributed quantities following some physical laws can optimally be extracted from data. IFT exploits known or inferred correlation structures of the field of interest s (s is regarded as a function $s = s(x)$ and as a vector $s = (s_x)_{x \in \Omega}$ in a function space) over some domain $\Omega = \{x\}$ in order to regularize the otherwise ill-posed inverse problem of determining virtually infinitely many field degrees of freedom from a finite dataset $d = (d_1, \dots, d_n)^T = (d_i)_i$, $n \in \mathbb{N}$. What distinguishes it from many non-parametric inference methods is that an IFT is defined over continuous spaces, and any pixelization of the field used in actual computations must preserve this continuum limit and recover it for infinitely small pixels. A concise introduction into IFT can be found Ref. [1], exhaustive ones in Refs. [2, 3], and the numerical issues of properly discretized fields are addressed in Refs. [4, 5].

Signal field estimation in IFT relies on Bayes theorem that gets recast into a (statistical) field theoretical language:

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)\mathcal{P}(s)}{\mathcal{P}(d)} \equiv \frac{e^{-\mathcal{H}(d,s)}}{\mathcal{Z}_d}. \quad (1)$$

This defines the information Hamiltonian and its partition function to be

$$\mathcal{H}(d,s) \equiv -\ln \mathcal{P}(d,s) = -\ln \mathcal{P}(d|s) - \ln \mathcal{P}(s) \text{ and} \quad (2)$$

$$\mathcal{Z}_d \equiv \int \mathcal{D}s e^{-\mathcal{H}(d,s)} = \int \mathcal{D}s \mathcal{P}(d,s) = \mathcal{P}(d), \quad (3)$$

and permits the usage of any appropriate field theoretical technique in order to do our signal inference. These include, among others, classical solutions (or Maximum A Posteriori (MAP) approximation), Feynman diagrams, renormalization and re-summation techniques, Gibbs free energy minimization (or Maximum Entropy), and mean field approaches. See Refs. [1, 3] for an overview. IFT can be applied in scientific areas that rely on imaging, like astronomy and astrophysics, and examples of such applications can be found at <https://www.mpa-garching.mpg.de/ift>. Useful IFT approaches to some generic astrophysical signal inference problems are reviewed below.

THE MEASUREMENT PROBLEM

In a generic linear measurement the data depend linearly on the signal,

$$d = R s + n = \left(\int_{\Omega} dx R_{ix} s_x + n_i \right)_i, \quad (4)$$

with R the instrument response operator, which describes how the instrument senses the signal, and n the noise. In the simplest case, signal and noise priors are independent Gaussian distributions,

$$\mathcal{P}(n, s) = \mathcal{G}(s, S) \mathcal{G}(n, N), \text{ with } \mathcal{G}(\phi, \Phi) = \frac{1}{\sqrt{|2\pi\Phi|}} \exp\left(-\frac{1}{2} \phi^\dagger \Phi^{-1} \phi\right), \quad (5)$$

with S and N known signal and noise covariances, respectively, and \dagger denoting the adjoint of a vector or field. Then, the signal posterior is Gaussian as well,

$$\mathcal{P}(s|d) = \mathcal{G}(s - m, D), \text{ with } m = D j, D = (S^{-1} + R^\dagger N^{-1} R)^{-1}, \text{ and } j = R^\dagger N^{-1} d. \quad (6)$$

Here, m is the signal posterior mean or Wiener filter estimate, D the signal uncertainty covariance or information propagator, and j the information source [3].

A typical signal in astronomy is the intensity distribution I_x of light of a certain wavelength on the celestial sphere $\Omega = \mathcal{S}^2$. Since I_x is strictly positive and can vary over orders of magnitude, the logarithmic brightness $s_x = \log(I_x/I_0)$ would be a natural variable that can be as well positive as negative. The *Maximum Entropy* principle singles out a Gaussian prior for s if only information up to the second moment of the s -distribution is available a priori or should be taken into account. Furthermore, the likelihood might be a Poisson distribution, if individual photons are counted. All this complicates the inference problem and introduces so-called interaction terms into the information Hamiltonian, but can in principle be dealt with the toolbox of IFT [3, 6, 7]. For simplicity, we assume in the following the measurement Eq. (4).

In practical applications of IFT, it is often the case that the signal covariance S , the noise covariance N , the response R , or combinations thereof are not sufficiently known for applying the Wiener filter given by Eq. (6). We denote all such undetermined parameters with the parameter vector p . Its uncertainty has to be taken into account in the signal inference. A p -marginalized joint probability of data and signal might be calculated analytically,

$$\mathcal{P}(d, s) = \int \mathcal{D}p \mathcal{P}(d, s, p). \quad (7)$$

However, the corresponding information Hamiltonian $\mathcal{H}(d, s) = -\ln \mathcal{P}(d, s)$ is usually a complicated function of the signal, preventing an easy calculation of the posterior mean m . Simple MAP signal estimators, resulting from solving $\partial \mathcal{H}(d, s) / \partial s = 0$ for s , can be highly biased due to the strong skewness of the marginalized posterior [8].

The *Empirical Bayes* approach performs much better. There, the s -marginalized parameter posterior $\mathcal{P}(d, p) = \int \mathcal{D}s \mathcal{P}(d, s, p)$ is used to choose a good point estimate p^* , which then is assumed to be correct in a signal mean inference based on $\mathcal{P}(s|d, p^*)$. The reason for *Empirical Bayes* to be a logically consistent approximation can be observed in the following reformulation of the posterior mean signal,

$$m \equiv \langle s \rangle_{(s,p|d)} \equiv \int \mathcal{D}p \int \mathcal{D}s \mathcal{P}(s, p|d) s = \int \mathcal{D}p \mathcal{P}(p|d) \int \mathcal{D}s \mathcal{P}(s|d, p) s = \langle \langle s \rangle_{(s|d,p)} \rangle_{(p|d)} \approx \langle s \rangle_{(s|d,p^*)}. \quad (8)$$

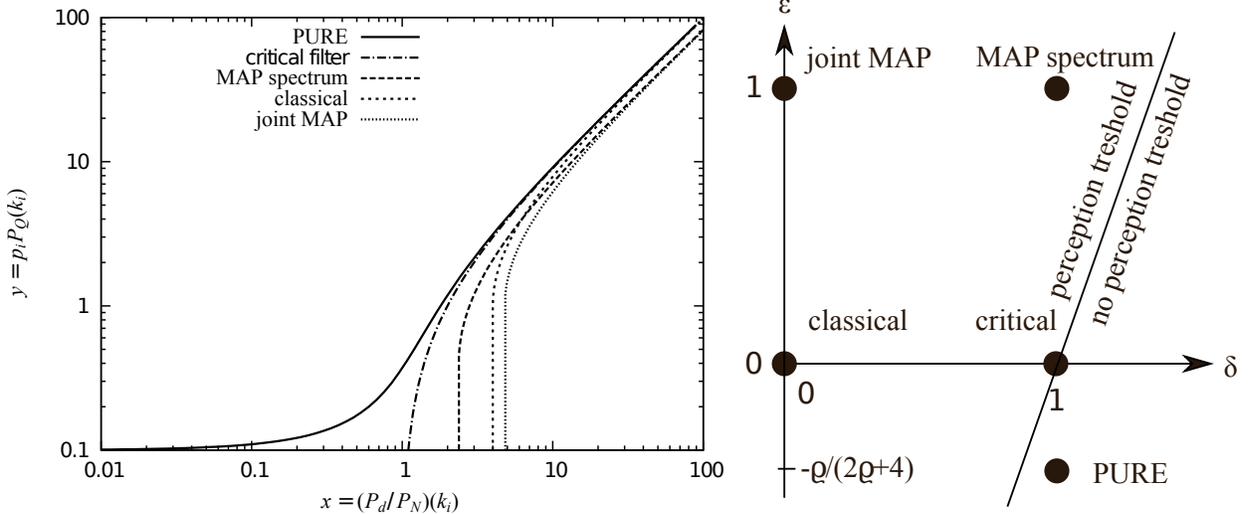
In the last step a delta-approximation, $\mathcal{P}(p|d) \approx \delta(p - p^*)$, was used. Thus, *Empirical Bayes* can be regarded as the zeroth order term of a perturbation expansion in the posterior parameter uncertainty. The next order term might be obtained by using a Gaussian approximation of $\mathcal{P}(p|d)$. A more systematic expansion can be obtained using the minimal Gibbs free energy approach [6]. However, already the empirical Bayes approximation leads to well-working signal inference recipes in case of unknown hyper-parameters for the problems of unknown covariances and/or response.

UNKNOWN SIGNAL COVARIANCE

The signal covariance S might be unknown, however, statistical homogeneity can often be assumed, i.e., the correlation $\langle s_x s_y \rangle_{(s)} = S_{xy} = C_s(x - y)$ is only a function of the distance between x , and y . This implies that S is diagonal in Fourier space,

$$S_{kk'} = (2\pi)^u \delta(k - k') P_s(k), \quad (9)$$

with k, k' being u -dimensional Fourier wave-vectors. In this case, a suitable hyper-prior for the unknown signal power spectrum $P_s(k)$ might be given by a term penalizing deviations from power-law spectra and inverse Gamma priors for



Source: Figures from [8].

the individual Fourier modes,

$$\mathcal{P}(P_s) \propto \exp \left[\underbrace{-\frac{1}{2} \int dk \left(\frac{\partial \log P_s(k)}{\sigma_k \partial \log k} \right)^2}_{\equiv -\frac{1}{2} \log P_s^\dagger T \log P_s} \right] \prod_k [P_s(k)]^{-\alpha_k} \exp \left(-\frac{q_k}{P_s(k)} \right). \quad (10)$$

The hyper-parameters σ_k , α_k , and q_k determine how strongly spectral curvature is penalized and how informative the spectral prior is for the individual modes. $\alpha_k = 1$ and $q_k = 0$ correspond to Jeffreys' non-informative prior. For the *Empirical Bayes* approximation we have to deduce a point estimate of the spectrum. Using the MAP approximation in $\log P_s(k)$ for this [7], we obtain

$$P_s^*(k) = \frac{q_k + \frac{1}{2} \text{Tr} \left[(m^* m^{*\dagger} + D^*) \mathbb{1}^{(k)} \right]}{\alpha_k - 1 + \frac{1}{2} \rho_k + (T \log P_s^*)_k}, \quad (11)$$

with $\mathbb{1}^{(k)}$ a projection onto all Fourier modes with the same power-spectrum as k (e.g., on spheres in Fourier space, in case of statistical isotropy) and $\rho_k = \text{Tr}[\mathbb{1}_k]$ the number of such modes. m^* and D^* are the mean and uncertainty covariance, respectively, calculated for the Gaussian inference problem assuming $P_s^*(k)$ to be self-consistently the correct power spectrum.

It is instructive to investigate this signal filter operation (the self-consistent calculation of m^* and P_s^*) in the context of similar filters. For this, we simplify the problem by assuming that signal and noise covariances, as well as response are diagonal in Fourier space, and that our spectral prior is completely non-informative ($\sigma_k = \infty$, $\alpha_k = 1$, and $q_k = 0$ for all k). A generic spectral estimator (to be calculated consistently with the corresponding signal estimator $m^* = D^* j$) is then

$$P_s^*(k) = \frac{\text{Tr} \left[(m^* m^{*\dagger} + \delta D^*) \mathbb{1}^{(k)} \right]}{\rho_k + \varepsilon}, \quad (12)$$

where the parameters δ and ε have been introduced in order to see how differently signal and spectrum estimators perform. The above derived non-informative filter corresponds to $(\delta, \varepsilon) = (1, 0)$, but other assumptions or approximations might lead to different values. These are depicted in the top right figure on this page. For example, the MAP estimator using the power spectrum marginalized posterior leads to $(\delta, \varepsilon) = (0, 0)$ (labeled “classical”) and a parameter uncertainty renormalized estimator (PURE) can be projected to $(\delta, \varepsilon) = (1, -\rho/(2\rho + 4))$. The above estimator with $(\delta, \varepsilon) = (1, 0)$ is called *critical filter*, because it resides on a critical line for perception thresholds. All the estimators to the left of this line exhibit a perception threshold. They do not respond at all to data modes with power below some critical value. This is depicted in the left figure, where the power in the resulting reconstructions of these filters is shown as a function of the ratio of data to noise power. Filters with perception threshold require the data

to have more variance than expected from the noise. For example, the *classical filter* with $(\delta, \varepsilon) = (0, 0)$ does not account for the missing power in the reconstruction. As a consequence it has a strong perception threshold. It requires the data to have four times the variance of the noise before it recognizes a signal. The *critical filter* has only a marginal perception threshold at the point where the data variance is exactly the noise variance. The PURE filter tries even to extract information from data with less variance than expected from the noise level.

UNKNOWN NOISE COVARIANCE

The critical filter needs to know precisely the noise level of the data in order to tune itself optimally. However, in many measurement situations this is also not reliably known and the data have to provide this information as well. As noise and signal variance are somehow degenerately encoded in the data, an informative prior for the noise level is mandatory whenever a non-informative prior for the signal power spectrum is used. The noise covariance might be decomposed as $N = \sum_i \eta_i N_i$, where the N_i denote block matrices in data space that contain the best available guesses for the error covariances. We model our noise prior knowledge with a multivariate inverse-gamma prior,

$$\mathcal{P}(\eta) \propto \prod_i \eta_i^{-\beta_i} \exp\left(-\frac{r_i}{\eta_i}\right), \quad (13)$$

by choosing $\beta_i > 1$ and $r_i > 0$ to represent our confidence in the reported or assumed error covariances N_i appropriately. The resulting signal estimation scheme is as before, just with an additional estimator for the noise parameters, given by

$$\eta_i^* = \frac{r_i + \frac{1}{2} \text{Tr} \left\{ [(d - Rm^*)(d - Rm^*)^\dagger + RD^*R^\dagger] N_i^{-1} \right\}}{\beta_i - 1 + \frac{1}{2} \mu_i}, \quad (14)$$

where N_i^{-1} is the pseudo-inverse of N_i and $\mu_i = \text{Tr}(N_i N_i^{-1})$ the dimensionality of the noise blocks [9]. This *extended critical filter* has successfully been used to reconstruct an all sky map of the galactic magnetic field [10].

UNKNOWN CALIBRATION

Finally, it might also well be that the response in Eq. (4) is not fully known. The process and the result of determining the response is called *calibration*. Ideally, the measurement of a well-known external calibration signal is used. However, often this is not sufficient, since the instrument response might change with time. In this case, the scientific signal of interest might be used as a calibrator itself. The resulting self-calibration scheme (*selfcal*) assumes some calibration, infers the signal with a reconstruction method, assumes this signal to be correct in order to calibrate on this, and iterates until convergence. Although the *selfcal* scheme is widely used, in particular in radio astronomy, an information theoretical investigation and a proof about its convergence was only provided recently [11].

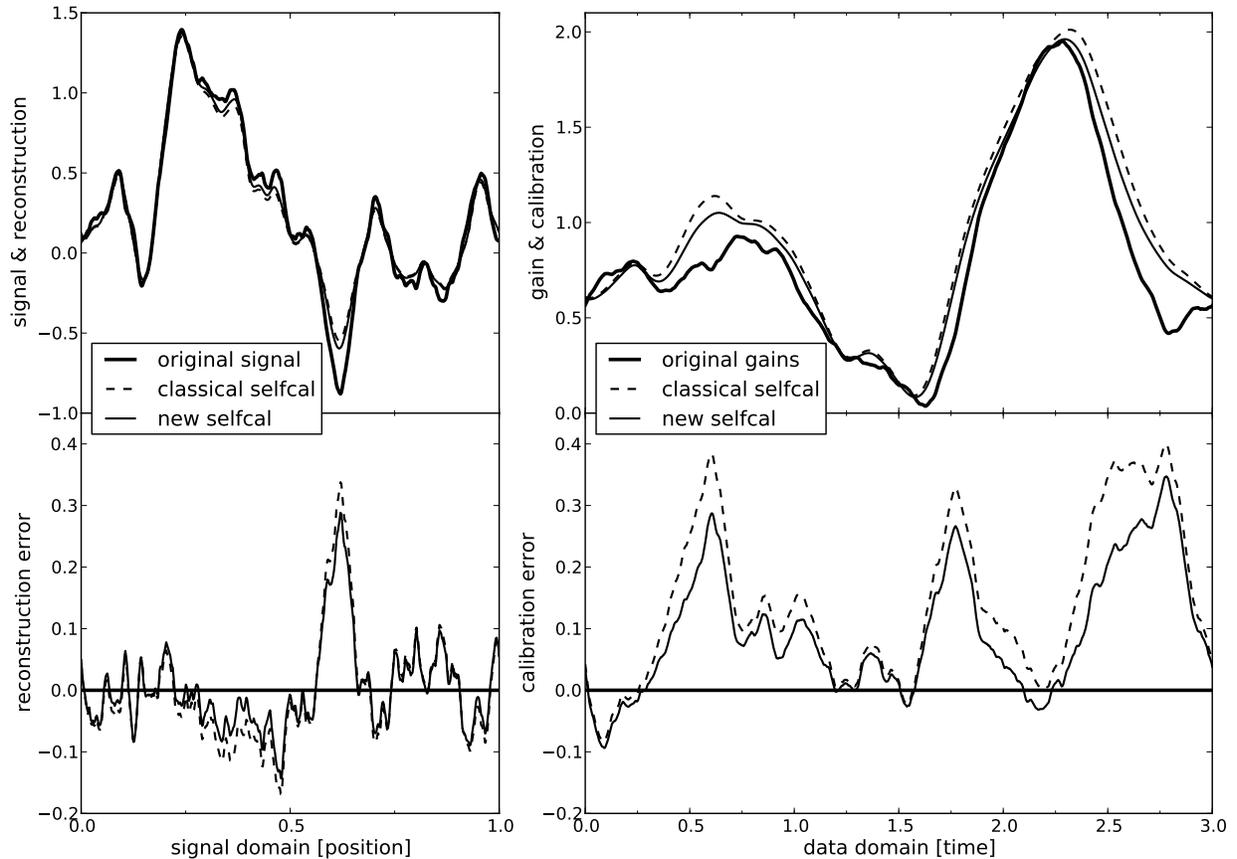
There it has been shown that if the imaging and calibration algorithms used can be regarded as MAP estimators, the classical *selfcal* scheme corresponds to a joint MAP estimation of signal and calibration. Similar to the case of unknown signal covariance, we expect such a joint MAP estimation to be sub-optimal in an \mathcal{L}^2 -error norm sense.

Specifically, we assume the unknown calibration parameters $\gamma = (\gamma_i)_i$ to affect the response linearly,

$$R^\gamma = B^0 + \sum_a \gamma_a B^a, \quad (15)$$

with B^0 and B^a known and γ -independent, and $\mathcal{P}(\gamma) = \mathcal{G}(\gamma, \Gamma)$ the calibration prior with some known calibration uncertainty matrix Γ . Using the *Empirical Bayes* approach, we find that the signal reconstruction m^* should use the calibration point estimate given by Ref. [11] self-consistently

$$\begin{aligned} \gamma^* &= \Delta^* h^*, \text{ with} \\ \Delta_{ab}^{*-1} &= \Gamma_{ab}^{-1} + \text{Tr} \left[(m^* m^{*\dagger} + D^*) B^{a\dagger} N^{-1} B^b \right], \text{ and} \\ h_b^* &= m^{*\dagger} B^{b\dagger} N^{-1} d - \text{Tr} \left[(m^* m^{*\dagger} + D^*) B^{0\dagger} N^{-1} B^b \right]. \end{aligned} \quad (16)$$



Source: Figures from [11].

This is a new *selfcal* scheme, which improves over *classical selfcal* by taking the posterior signal uncertainty $D^* = \langle (s - m^*)(s - m^*)^\dagger \rangle_{(s|d, \gamma^*)}$ into account.

An example of *selfcal* can be seen in the figure on this page. There, an unknown signal is shown, which was scanned three times with an instrument with varying gains (also shown) according to the measurement equation $d_t = (1 + \gamma_t) s_{x_t} + n_t$ with $x_t = t \bmod 1$. The solutions of the *classical selfcal* and the *new selfcal* schemes are compared to this. Both solve Eq. (6) and Eq. (16) simultaneously, while *classical selfcal* ignores D^* in Eq. (16). The figure shows that the *new selfcal* scheme alleviates partly a bias of *classical selfcal* for high gain solutions.

ASTROPHYSICAL APPLICATIONS

The versatility of IFT to deal with various measurement problems, to take uncertainties on covariances and responses into account and to provide implementable algorithms have let to a number of concrete IFT applications in astronomy and astrophysics. Many of them build on NIFTY, Numerical Information Field Theory [4, 5], a publicly available, object-oriented library that permits the user to program IFT algorithms abstractly, irrespective of the space discretization used. IFT was already applied to a number of areas, which we briefly discuss in the following.

The **cosmic microwave background** statistic is nearly Gaussian. However, different inflationary scenarios of cosmology predict different levels of non-Gaussian signatures. Optimal non-Gaussianity estimators could be constructed using Feynman diagrams [3] and by a saddle point approximation [12].

The **cosmic large-scale structure** of the matter distribution in the Universe is traced by galaxies and can therefore be reconstructed from galaxy surveys [13, 14, 15, and others]. The power spectrum of the large-scale structure can be measured as well [14] and analyzed for its information content on cosmology. Although the cosmic density field is a strictly positive quantity, it is often treated as a Gaussian random field. A log-normal model would be more accurate

[3, 16, 17]. For its usage, it is necessary to translate between linear and logarithmic spectra [18].

Galactic magnetism can be studied using the Faraday rotation of extragalactic radio sources. The construction of all-sky Faraday maps from such individual measurements towards the directions of these sources was possible thanks to the *critical filter* [19] and the *extended critical filter* [20].

Interferometric radio astronomy faces the problem that an interferometer array measures only some part of the Fourier transformed sky brightness. The other part has to be reconstructed by the imaging algorithm. *RESOLVE*, a novel IFT based imaging method, which assumes the sky brightness to follow a log-normal distribution with unknown power spectrum, seems to be superior in reconstructing diffuse emission compared to classical algorithms (e.g. *CLEAN* and *MAXENT*) [21].

Gamma- and X-ray astronomy have to deal with extended emission, point sources, Poisson statistics, inhomogeneous sky exposure, and complex point spread functions. The IFT based D³PO algorithm for *denoising, deconvolving, and decomposition of photon observations* handles these and produces maps of the extended emission, catalogs of the point sources, and angular power spectra of the diffuse flux [22].

These applications demonstrate that IFT is a versatile framework to develop and analyze measurement and imaging problems not only in astronomy, but also in cosmology and other areas.

ACKNOWLEDGMENTS

First of all, I should apologize for exclusively reviewing IFT work I was involved in. There is a large amount of related work I had to ignore for brevity. I want to thank my IFT co-investigators of the here discussed works, Michael Bell, Sebastian Dorn, Mona Frommert, Maksim Greiner, Jens Jasche, Henrik Junklewitz, Righi Khatri, Francisco Shu Kitaura, Niels Oppermann, Martin Reinecke, Georg Robbers, Marco Selig, and Cornelius Weig. Furthermore, I acknowledge helpful comments on the manuscript by Vanessa Böhm, Sebastian Dorn, and Marco Selig.

REFERENCES

1. T. Enßlin, “Information field theory,” in *American Institute of Physics Conference Series*, edited by U. von Toussaint, 2013, vol. 1553 of *American Institute of Physics Conference Series*, pp. 184–191, 1301.2556.
2. J. C. Lemm, *Bayesian Field Theory*, Johns Hopkins University Press, 2003.
3. T. A. Enßlin, M. Frommert, and F. S. Kitaura, *Phys.Rev.D* **80**, 105005 (2009), 0806.3474.
4. M. Selig, M. R. Bell, H. Junklewitz, N. Oppermann, M. Reinecke, M. Greiner, C. Pachajoa, and T. A. Enßlin, *A&A* **554**, A26 (2013), 1301.4499.
5. M. Selig, “The NIFTY way of Bayesian signal inference,” in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, 2013, vol. these proceedings.
6. T. A. Enßlin, and C. Weig, *Phys.Rev.E* **82**, 051112 (2010), 1004.2868.
7. N. Oppermann, M. Selig, M. R. Bell, and T. A. Enßlin, *Phys.Rev.E* **87**, 032136 (2013), 1210.6866.
8. T. A. Enßlin, and M. Frommert, *Phys.Rev.D* **83**, 105014 (2011), 1002.2928.
9. N. Oppermann, G. Robbers, and T. A. Enßlin, *Phys.Rev.E* **84**, 041118 (2011), 1107.2384.
10. N. Oppermann et al., *A&A* **542**, A93 (2012), 1111.6186.
11. T. A. Enßlin, H. Junklewitz, L. Winderling, and M. Selig, *ArXiv e-prints* (2013), 1312.1349.
12. S. Dorn, N. Oppermann, R. Khatri, M. Selig, and T. A. Enßlin, *Phys.Rev.D* **88**, 103516 (2013), 1307.3884.
13. F. S. Kitaura, and T. A. Enßlin, *MNRAS* **389**, 497–544 (2008), 0705.0429.
14. J. Jasche, F. S. Kitaura, B. D. Wandelt, and T. A. Enßlin, *MNRAS* **406**, 60–85 (2010), 0911.2493.
15. F.-S. Kitaura, P. Erdoğdu, S. E. Nuza, A. Khalatyan, R. E. Angulo, Y. Hoffman, and S. Gottlöber, *MNRAS* **427**, L35–L39 (2012), 1205.5560.
16. F. S. Kitaura et al., *MNRAS* **400**, 183–203 (2009), 0906.3978.
17. C. Weig, and T. A. Enßlin, *MNRAS* **409**, 1393–1411 (2010), 1003.1311.
18. M. Greiner, and T. A. Enßlin, *ArXiv e-prints* (2013), 1312.1354.
19. N. Oppermann, H. Junklewitz, G. Robbers, and T. A. Enßlin, *A&A* **530**, A89+ (2011), 1008.1246.
20. N. Oppermann, H. Junklewitz, G. Robbers, M. R. Bell, T. A. Enßlin, A. Bonafede, R. Braun, J. C. Brown, T. E. Clarke, I. J. Feain, B. M. Gaensler, A. Hammond, L. Harvey-Smith, G. Heald, M. Johnston-Hollitt, U. Klein, P. P. Kronberg, S. A. Mao, N. M. McClure-Griffiths, S. P. O’Sullivan, L. Pratley, T. Robishaw, S. Roy, D. H. F. M. Schnitzeler, C. Sotomayor-Beltran, J. Stevens, J. M. Stil, C. Sunstrum, A. Tanna, A. R. Taylor, and C. L. Van Eck, *A&A* **542**, A93 (2012), 1111.6186.
21. H. Junklewitz, M. R. Bell, M. Selig, and T. A. Enßlin, *ArXiv e-prints* (2013), 1311.5282.
22. M. Selig, and T. Enßlin, *ArXiv e-prints* (2013), 1311.1888.