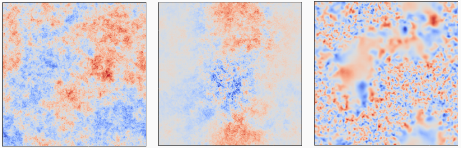
Gaussian & Generative processes

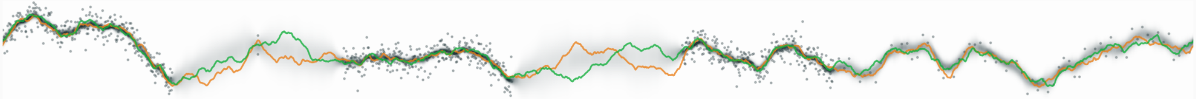


Encoding prior knowledge for field inference problems

Philipp Frank[1]

IMAGINE workshop: Towards a comprehensive model of the galactic magnetic field, NORDITA, Stockholm, April 11, 2023

(1) Max-Planck Institute for Astrophysics MPA, Garching, Germany



Gaussian processes

- Probability distributions $\mathcal{P}(s)$ over functions $s_x \equiv s(x)$, with $s \in \mathcal{L}^2[\Omega]$, $x \in \Omega \subset \mathbb{R}^N$.

Gaussian processes

- Probability distributions $\mathcal{P}(s)$ over functions $s_x \equiv s(x)$, with $s \in \mathcal{L}^2[\Omega]$, $x \in \Omega \subset \mathbb{R}^N$.
- Gaussian processes are fully specified by their one- and two-point correlation functions:
- Mean field: $m_x = m(x) \equiv \langle s_x \rangle_{\mathcal{P}(s)}$.
- Correlation structure: $C_{xy} = C(x, y) \equiv \left\langle (s_x - m_x)(s_y - m_y)^* \right\rangle_{\mathcal{P}(s)}$.

Gaussian processes

- Probability distributions $\mathcal{P}(s)$ over functions $s_x \equiv s(x)$, with $s \in \mathcal{L}^2[\Omega]$, $x \in \Omega \subset \mathbb{R}^N$.
- Gaussian processes are fully specified by their one- and two-point correlation functions:
- Mean field: $m_x = m(x) \equiv \langle s_x \rangle_{\mathcal{P}(s)}$.
- Correlation structure: $C_{xy} = C(x, y) \equiv \left\langle (s_x - m_x)(s_y - m_y)^* \right\rangle_{\mathcal{P}(s)}$.

Gaussian process (GP) distribution

$$\begin{aligned}\mathcal{P}(s|m, C) = \mathcal{N}(s; m, C) &= \frac{1}{\sqrt{|2\pi C|}} \exp\left(-\frac{1}{2}(s-m)^\dagger C^{-1}(s-m)\right)\\ &= \frac{1}{\sqrt{|2\pi C|}} \exp\left(-\frac{1}{2}\int\int (s(x)-m(x))\; C^{-1}(x,y)\; (s(y)-m(y))\,\mathrm{d}x\mathrm{d}y\right)\end{aligned}$$

Inverse operator: C^{-1}; Functional determinant: $|\bullet|$.

Gaussian processes - Regression

Gaussian distributions are closed under linear maps: Every linear transformation of a Gaussian is also Gaussian.

Gaussian processes - Regression

Gaussian distributions are closed under linear maps: Every linear transformation of a Gaussian is also Gaussian.

- Every discretization $(\mathbf{s})_i \equiv s(x_i)$, with $x_i \in \Omega\ \forall i \in \{1, ..., N\}$ is distributed according to a multivariate Gaussian:
- With mean $\mathbf{m}_i = m(x_i)$ and Covariance matrix $\mathbf{C}_{ij} = C(x_i, x_j)$.

Multivariate Gaussian

$$\mathcal{P}(\mathbf{s}|m, C) = \mathcal{N}(\mathbf{s}; \mathbf{m}, \mathbf{C}) = \frac{1}{\sqrt{|2\pi\mathbf{C}|}} \exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{m})^\dagger \mathbf{C}^{-1}(\mathbf{s}-\mathbf{m})\right)$$

Matrix inverse: $\mathbf{C}^{-1}$; Determinant: $|\bullet|$.

Gaussian processes - Regression

Gaussian distributions are closed under linear maps: Every linear transformation of a Gaussian is also Gaussian.

- Every discretization $(\mathbf{s})_i \equiv s(x_i)$, with $x_i \in \Omega\ \forall i \in \{1, ..., N\}$ is distributed according to a multivariate Gaussian:
- With mean $\mathbf{m}_i = m(x_i)$ and Covariance matrix $\mathbf{C}_{ij} = C(x_i, x_j)$.

Multivariate Gaussian

$$\mathcal{P}(\mathbf{s}|m, C) = \mathcal{N}(\mathbf{s}; \mathbf{m}, \mathbf{C}) = \frac{1}{\sqrt{|2\pi\mathbf{C}|}} \exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{m})^\dagger \mathbf{C}^{-1}(\mathbf{s}-\mathbf{m})\right)$$

Matrix inverse: $\mathbf{C}^{-1}$; Determinant: $|\bullet|$.

- Given a set of linear measurements $\mathbf{d}$ of $\mathbf{s}$, the resulting posterior $\mathcal{P}(\mathbf{s}|\mathbf{d}, m, C)$ is also Gaussian $\rightarrow$ *Linear filter theory*.

Gaussian processes - Regression

Gaussian distributions are closed under linear maps: Every linear transformation of a Gaussian is also Gaussian.

- Every discretization $(\mathbf{s})_i \equiv s(x_i)$, with $x_i \in \Omega\ \forall i \in \{1, ..., N\}$ is distributed according to a multivariate Gaussian:
- With mean $\mathbf{m}_i = m(x_i)$ and Covariance matrix $\mathbf{C}_{ij} = C(x_i, x_j)$.

Multivariate Gaussian

$$\mathcal{P}(\mathbf{s}|m, C) = \mathcal{N}(\mathbf{s}; \mathbf{m}, \mathbf{C}) = \frac{1}{\sqrt{|2\pi\mathbf{C}|}} \exp\left(-\frac{1}{2}(\mathbf{s}-\mathbf{m})^\dagger \mathbf{C}^{-1}(\mathbf{s}-\mathbf{m})\right)$$

Matrix inverse: $\mathbf{C}^{-1}$; Determinant: $|\bullet|$.

- Given a set of linear measurements $\mathbf{d}$ of $\mathbf{s}$, the resulting posterior $\mathcal{P}(\mathbf{s}|\mathbf{d}, m, C)$ is also Gaussian $\rightarrow$ *Linear filter theory*.
- Main computational challenges: Apply matrix inverse C^{-1}, construct conditionals (Schur complement), (for kernel estimation) evaluate determinant.
- Non-linear measurements? How to estimate C?

Generative (Gaussian) processes

- Given measurements d and a generic (non-Gaussian) Likelihood $\mathcal{P}(d|s)$,
- Together with a GP-prior $\mathcal{P}(s) = \mathcal{N}(s; 0, C)$.

Joint distribution

$$\mathcal{P}(s, d|C) = \mathcal{P}(d|s)\, \mathcal{N}(s; 0, C) \ .$$

Generative (Gaussian) processes

- Given measurements d and a generic (non-Gaussian) Likelihood $\mathcal{P}(d|s)$,
- Together with a GP-prior $\mathcal{P}(s) = \mathcal{N}(s; 0, C)$.
- Let $s_x = (A\ \xi)_x \equiv \int A(x, y)\ \xi(y)\ \mathrm{d}y$ with $C = AA^\dagger$
- Since $\mathcal{P}(s)\ \mathrm{d}s \overset{!}{=} \mathcal{P}(\xi)\ \mathrm{d}\xi$ we get $\mathcal{P}(\xi) = \mathcal{N}(\xi; 0, \mathbb{1})$.

Joint distribution

$$\mathcal{P}(s, d|C) = \mathcal{P}(d|s)\ \mathcal{N}(s; 0, C)\ .$$

Generative (Gaussian) processes

- Given measurements d and a generic (non-Gaussian) Likelihood $\mathcal{P}(d|s)$,
- Together with a GP-prior $\mathcal{P}(s) = \mathcal{N}(s; 0, C)$.
- Let $s_x = (A\,\xi)_x \equiv \int A(x,y)\,\xi(y)\,\mathrm{d}y$ with $C = AA^\dagger$
- Since $\mathcal{P}(s)\,\mathrm{d}s \stackrel{!}{=} \mathcal{P}(\xi)\,\mathrm{d}\xi$ we get $\mathcal{P}(\xi) = \mathcal{N}(\xi; 0, \mathbb{1})$.

Joint distribution

$$\mathcal{P}(\xi, d|A) = \mathcal{P}(d|s = A\,\xi)\,\mathcal{N}(\xi; 0, \mathbb{1})\ .$$

$$s = A\,\xi;$$

Generative (Gaussian) processes

- Given measurements d and a generic (non-Gaussian) Likelihood $\mathcal{P}(d|s)$,
- Together with a GP-prior $\mathcal{P}(s) = \mathcal{N}(s; 0, C)$.
- Let $s_x = (A\ \xi)_x \equiv \int A(x, y)\ \xi(y)\ \mathrm{d}y$ with $C = AA^\dagger$
- Since $\mathcal{P}(s)\ \mathrm{d}s \overset{!}{=} \mathcal{P}(\xi)\ \mathrm{d}\xi$ we get $\mathcal{P}(\xi) = \mathcal{N}(\xi; 0, \mathbb{1})$.

Joint distribution

$$\mathcal{P}(\xi, d|A) = \mathcal{P}(d|s = A\ \xi)\ \mathcal{N}(\xi; 0, \mathbb{1})\ .$$

$$s = A\ \xi;$$

- Joint inference of A (thereby also C): Let $A = A(\xi_A)$ be a generative model of A, with $\mathcal{P}(\xi_A; 0, \mathbb{1})$.

Generative (Gaussian) processes

- Given measurements d and a generic (non-Gaussian) Likelihood $\mathcal{P}(d|s)$,
- Together with a GP-prior $\mathcal{P}(s) = \mathcal{N}(s; 0, C)$.
- Let $s_x = (A\,\xi)_x \equiv \int A(x,y)\,\xi(y)\,\mathrm{d}y$ with $C = AA^\dagger$
- Since $\mathcal{P}(s)\,\mathrm{d}s \overset{!}{=} \mathcal{P}(\xi)\,\mathrm{d}\xi$ we get $\mathcal{P}(\xi) = \mathcal{N}(\xi; 0, \mathbb{1})$.

Joint distribution

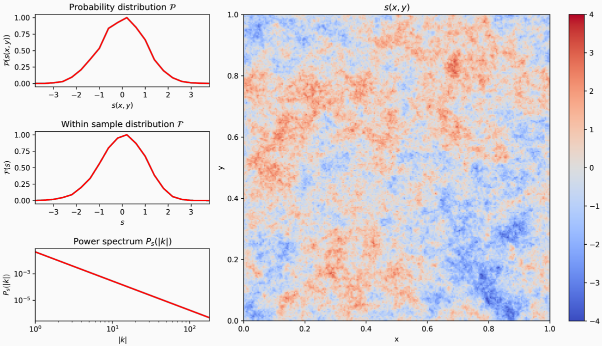
$$\mathcal{P}(\xi, \xi_A, d) = \mathcal{P}(d|s = A(\xi_A)\,\xi)\,\mathcal{N}(\xi; 0, \mathbb{1})\,\mathcal{N}(\xi_A; 0, \mathbb{1}).$$

$$s = A\,\xi;\; A = A(\xi_A).$$

- Joint inference of A (thereby also C): Let $A = A(\xi_A)$ be a generative model of A, with $\mathcal{P}(\xi_A; 0, \mathbb{1})$.

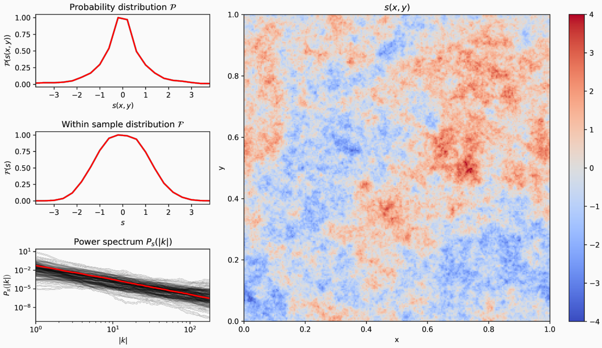
GP - Priors

$$s = A\,\xi, \quad \text{with} \quad A\left(\vec{x}, \vec{x'}\right) \propto \left|\vec{x} - \vec{x'}\right|^{-\alpha}.$$



GP - Priors

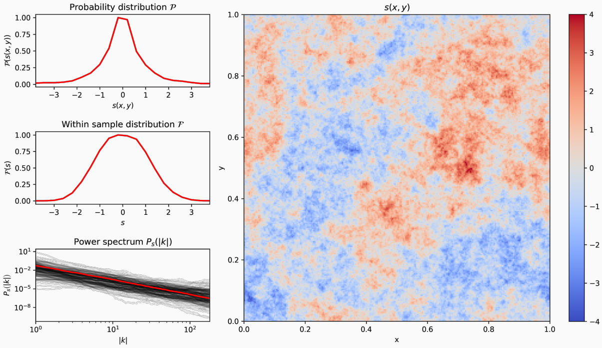
$$s = A\,\xi, \quad \text{with} \quad A \propto \mathcal{F}^{-1}\widehat{\sqrt{P_s}}, \quad P_s(k) \propto e^{\tau(k)}\ .$$



VLBI - M87* [AFH+22]

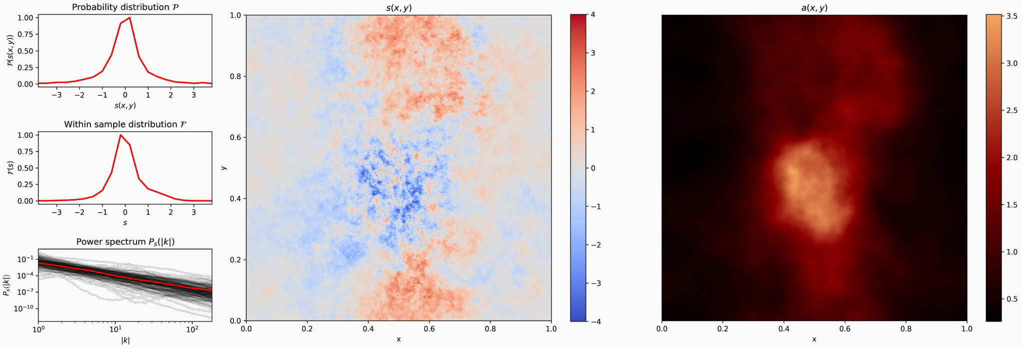
GP - Priors

$$s = A\,\xi, \quad \text{with} \quad A \propto \mathcal{F}^{-1}\widehat{\sqrt{P_s}}, \quad P_s(k) \propto e^{\tau(k)}\ .$$

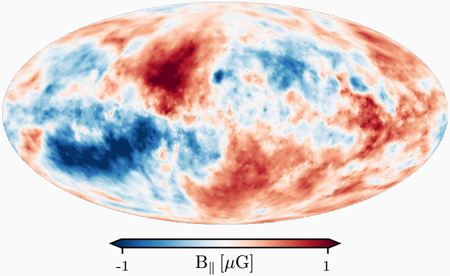


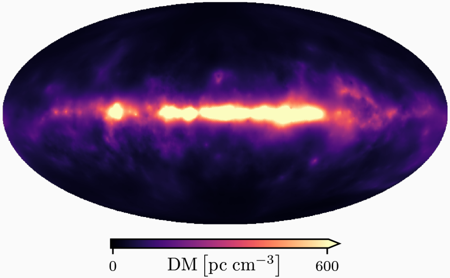
GP - Priors

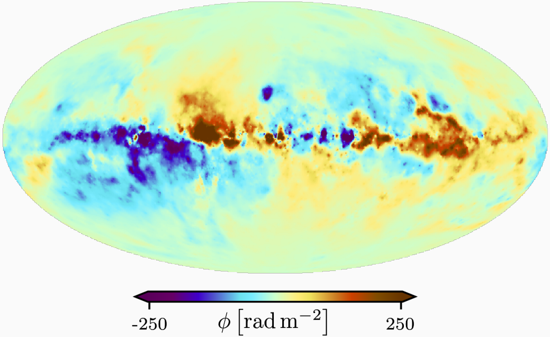
$$s = A\,\xi, \quad \text{with} \quad A \propto \hat{a}\,\mathcal{F}^{-1}\widehat{\sqrt{P_s}}, \quad P_s(k) \propto e^{\tau(k)}\ .$$



Applications [HE20]

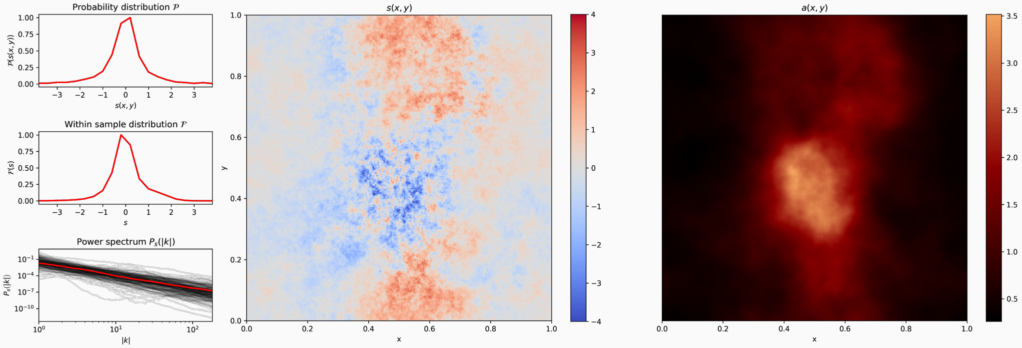






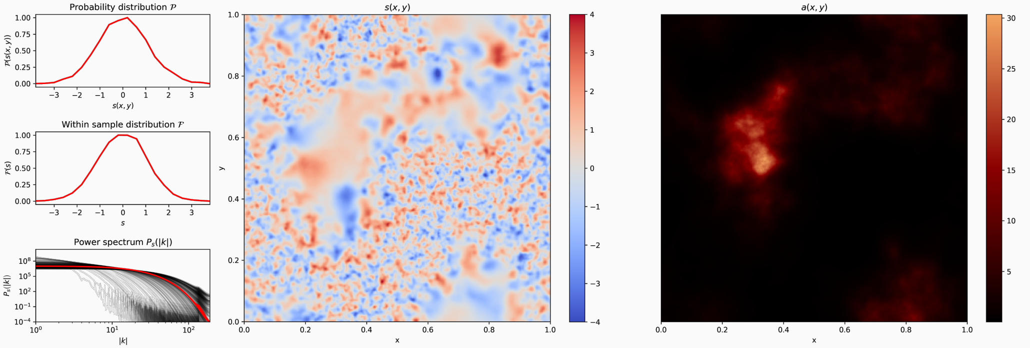
GP - Priors

$$s = A\,\xi, \quad \text{with} \quad A \propto \hat{a}\,\mathcal{F}^{-1}\widehat{\sqrt{P_s}}, \quad P_s(k) \propto e^{\tau(k)}\ .$$

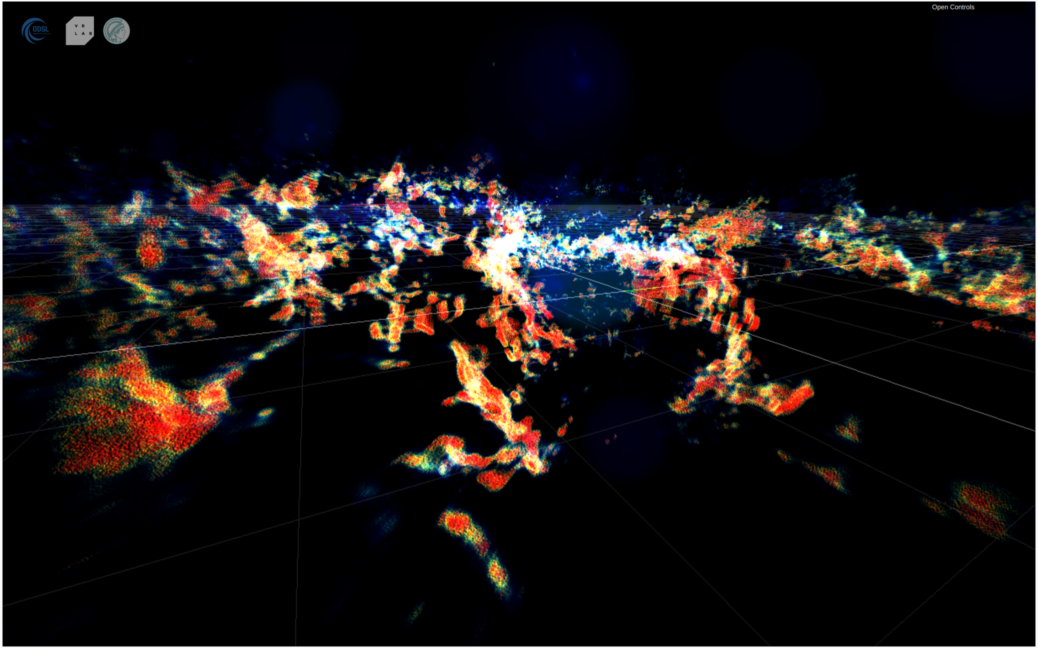


GP - Priors

$$s = A\,\xi, \quad \text{with} \quad A\left(\vec{x},\vec{x}'\right) \propto 1\Big/\left(1+\frac{1}{\sigma(a(\vec{x}))}\left|\vec{x}-\vec{x}'\right|^2\right)^2 .$$



Dust tomography [LEK+22]



GP - Summary & Outlook [ELFE22]

Generative process → Integral equation

$$s\left(\vec{x}\right) = \int A\left(\vec{x}, \vec{y}\right)\ \xi\left(\vec{y}\right)\ \mathrm{d}\vec{y}\ .$$

GP - Summary & Outlook [ELFE22]

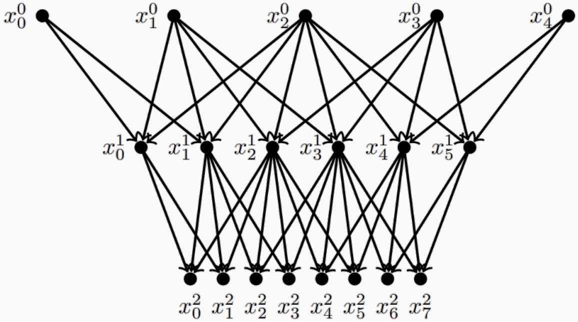
Generative process → Integral equation

$$s(\vec{x}) = \int A(\vec{x}, \vec{y})\ \xi(\vec{y})\ \mathrm{d}\vec{y} = \int A(\vec{x} - \vec{y})\ \xi(\vec{y})\ \mathrm{d}\vec{y}\ .$$

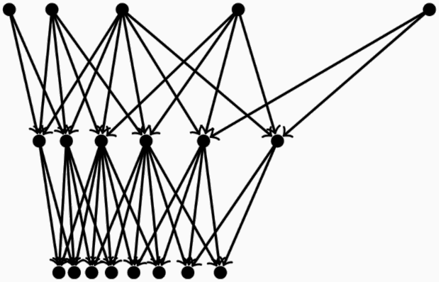
GP - Summary & Outlook [ELFE22]

Generative process → Integral equation

$$s(\vec{x}) = \int A(\vec{x}, \vec{y})\ \xi(\vec{y})\ \mathrm{d}\vec{y} = \int A\left(\vec{x}\left(\vec{\phi}\right) - \vec{y}\left(\vec{\phi'}\right)\right)\ \xi\left(\vec{y}\left(\vec{\phi'}\right)\right) \mathrm{d}\vec{\phi'}\ .$$



(a) Regular



(b) Logarithmic

GP - Conclusions

$$s(\vec{x}) = \int A(\vec{x}, \vec{y}) \;\xi(\vec{y}) \,\mathrm{d}\vec{y} \;.$$

- Gaussian Prior $\neq$ Gaussian Posterior
- Prior model $\leftrightarrow$ Physical model
- Solving inference problems given data: $\mathcal{O}\left(10^4 \sim 10^5\right)$ model evaluations.
- Prior vs. Likelihood: "natural" grids (e.g. Stat. homogeneity $\rightarrow$ Euclidean grid vs. Gal. Tomography $\rightarrow$ Radial grids)
- Calibration of prior models using simulations:
- Two-point correlation function of observables
- Identifying most relevant higher moments
- ...

Numerical Information Field Theory (NIFTy)

NIFTY

Code: https://gitlab.mpcdf.mpg.de/ift/nifty

Docs: https://ift.pages.mpcdf.de/nifty

- Philipp Arras, Philipp Frank, Philipp Haim, Jakob Knollmüller, Reimar Leike, Martin Reinecke, and Torsten A. Enßlin.
  Variable structures in m87* from space, time and frequency resolved interferometry.
  Nature Astronomy, 6(2):259–269, 2022.
- Gordian Edenhofer, Reimar H. Leike, Philipp Frank, and Torsten A. Enßlin.
  Sparse kernel gaussian processes through iterative charted refinement (icr), 2022.
- Sebastian Hutschenreuter and Torsten A. Enßlin.
  The galactic faraday depth sky revisited.
  A&A, 633:A150, 2020.

References ii

Reimar Leike, Gordian Edenhofer, Jakob Knollmüller, Christian Alig, Philipp Frank, and Torsten A. Enßlin.
The galactic 3d large-scale dust distribution via gaussian process regression on spherical coordinates.
arXiv, 2204.11715, 2022.

Leike, Reimar H., Glatze, Martin, and Enßlin, Torsten A.
Resolving nearby dust clouds.
A&A, 639:A138, 2020.