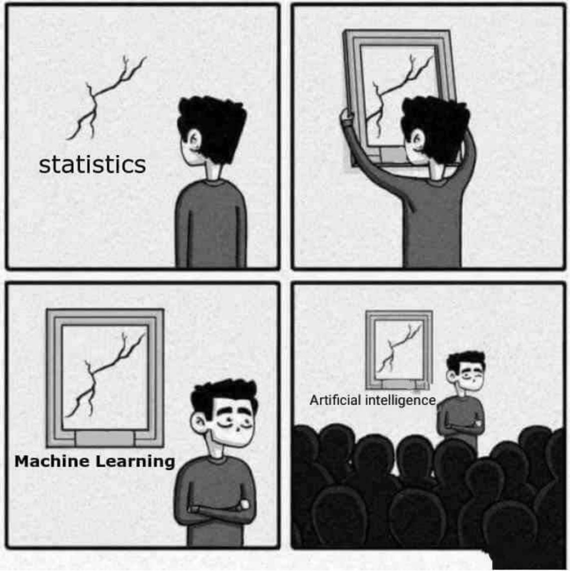
# Dynamical modelling of dwarf spheroidal galaxies using Gaussian-process emulation

MACHINE LEARNING APPLICATIONS FOR ASTRONOMY
UNIVERSITY OF NOTTINGHAM

**Amery Gration and Mark I. Wilkinson**

University of Leicester

11 September 2018


statistics
Machine Learning
Artificial intelligence

# Dwarf-spheroidal galaxies

**Figure:** The Fornax dwarf-spheroidal galaxy (ESO/Digitized Sky Survey 2).

# Dwarf-spheroidal galaxies

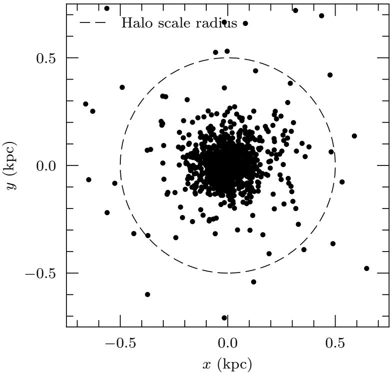



**Figure:** Nucleus of $10^5$ to $10^7$ stars within dark-matter halo of $10^9\,M_\odot$.

# The core-cusp problem

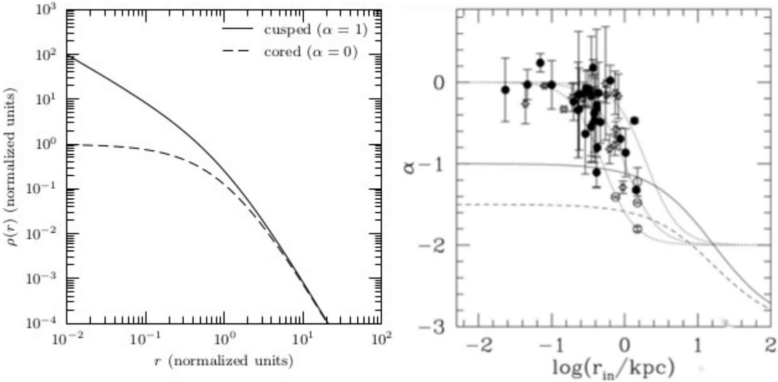



**Figure:** Left: cored and cusped density profiles. Right: observed values for the inner log slope (**Blok, W.J.G. de, *et al.*, 2001. ApJ.**)

# Distribution-function modelling

- Treat a star's state, $\boldsymbol{w} = (\boldsymbol{x}, \boldsymbol{v})$, as the realization of a random variable, with probability density function modelled by $(f_{\boldsymbol{W}}(\cdot\,; \boldsymbol{a}))_{\boldsymbol{a} \in \boldsymbol{A}}$, for model parameter $\boldsymbol{a}$, and parameter space $\boldsymbol{A}$.

# Distribution-function modelling

- Treat a star's state, $\boldsymbol{w} = (\boldsymbol{x}, \boldsymbol{v})$, as the realization of a random variable, with probability density function modelled by $(f_{\boldsymbol{W}}(\cdot; \boldsymbol{a}))_{\boldsymbol{a} \in \boldsymbol{A}}$, for model parameter $\boldsymbol{a}$, and parameter space $\boldsymbol{A}$.
- We observe the sky position and line-of-sight velocity, $\boldsymbol{w}_{\mathrm{p}} = (\boldsymbol{x}_{\mathrm{p}}, v_z)$:

$$f_{\boldsymbol{W}_{\mathrm{p}}}(\boldsymbol{w}_{\mathrm{p}}; \boldsymbol{a}) = \int_{\mathbf{R}} \int_{\mathbf{R}} \int_{\mathbf{R}} f_{\boldsymbol{W}}(\boldsymbol{w}; \boldsymbol{a}) \mathrm{d}z \mathrm{d}v_x, \mathrm{d}v_y.$$

# Distribution-function modelling

- Treat a star's state, $\boldsymbol{w} = (\boldsymbol{x}, \boldsymbol{v})$, as the realization of a random variable, with probability density function modelled by $(f_{\boldsymbol{W}}(\cdot; \boldsymbol{a}))_{\boldsymbol{a} \in \boldsymbol{A}}$, for model parameter $\boldsymbol{a}$, and parameter space $\boldsymbol{A}$.
- We observe the sky position and line-of-sight velocity, $\boldsymbol{w}_\mathrm{p} = (\boldsymbol{x}_\mathrm{p}, v_z)$:

$$f_{\boldsymbol{W}_\mathrm{p}}(\boldsymbol{w}_\mathrm{p}; \boldsymbol{a}) = \int_{\mathbf{R}} \int_{\mathbf{R}} \int_{\mathbf{R}} f_{\boldsymbol{W}}(\boldsymbol{w}; \boldsymbol{a}) \mathrm{d}z \mathrm{d}v_x, \mathrm{d}v_y.$$

- For $N$ stars the joint probability density function is

$$f(\boldsymbol{w}_{\mathrm{p},1}, \boldsymbol{w}_{\mathrm{p},2}, \ldots, \boldsymbol{w}_{\mathrm{p},N}; \boldsymbol{a}) = \prod_{i=1}^{N} f_{\boldsymbol{W}_{\mathrm{p},i}}(\boldsymbol{w}_{\mathrm{p,i}}; \boldsymbol{a}).$$

# Distribution-function modelling

- Treat a star's state, $\boldsymbol{w} = (\boldsymbol{x}, \boldsymbol{v})$, as the realization of a random variable, with probability density function modelled by $(f_{\boldsymbol{W}}(\cdot; \boldsymbol{a}))_{\boldsymbol{a} \in \boldsymbol{A}}$, for model parameter $\boldsymbol{a}$, and parameter space $\boldsymbol{A}$.
- We observe the sky position and line-of-sight velocity, $\boldsymbol{w}_{\mathrm{p}} = (\boldsymbol{x}_{\mathrm{p}}, v_z)$:

$$f_{\boldsymbol{W}_{\mathrm{p}}}(\boldsymbol{w}_{\mathrm{p}}; \boldsymbol{a}) = \int_{\mathbf{R}} \int_{\mathbf{R}} \int_{\mathbf{R}} f_{\boldsymbol{W}}(\boldsymbol{w}; \boldsymbol{a}) \mathrm{d}z \mathrm{d}v_x, \mathrm{d}v_y.$$

- For $N$ stars the joint probability density function is

$$f(\boldsymbol{w}_{\mathrm{p},1}, \boldsymbol{w}_{\mathrm{p},2}, \ldots, \boldsymbol{w}_{\mathrm{p},N}; \boldsymbol{a}) = \prod_{i=1}^{N} f_{\boldsymbol{W}_{\mathrm{p},i}}(\boldsymbol{w}_{\mathrm{p,i}}; \boldsymbol{a}).$$

- Recover $\boldsymbol{a}$ using the likelihood, $L(\boldsymbol{a}) := f(\boldsymbol{w}_{\mathrm{p},1}, \boldsymbol{w}_{\mathrm{p},2}, \ldots, \boldsymbol{w}_{\mathrm{p},N}; \boldsymbol{a})$.

# Toy model: anisotropic Plummer sphere

- Potential-density pair due to Plummer (1911):

$$\Phi(r) = -\frac{GM}{\sqrt{r^2 + b^2}}, \text{ and } \rho(r) = \frac{3M}{4\pi b^3}\left(1 + \frac{r^2}{b^2}\right)^{-5/2}.$$

# Toy model: anisotropic Plummer sphere

- Potential-density pair due to Plummer (1911):

$$\Phi(r) = -\frac{GM}{\sqrt{r^2 + b^2}}, \text{ and } \rho(r) = \frac{3M}{4\pi b^3}\left(1 + \frac{r^2}{b^2}\right)^{-5/2}.$$

- Using Osipkov and Merritt's method (Osipkov, 1975, and Merritt, 1985) we find the DF is

$$f_Q(Q) = \frac{3Mb^2}{\pi^3\sqrt{2}r_{\mathrm{a}}^2}\left(\frac{16(r_{\mathrm{a}}^2 - b^2)}{7}Q^{7/2} + (GM)^2 Q^{3/2}\right).$$

where

$$Q = -E - \frac{L^2}{2r_a}.$$

# Toy model: anisotropic Plummer sphere

- Potential-density pair due to Plummer (1911):

$$\Phi(r) = -\frac{GM}{\sqrt{r^2 + b^2}}, \text{ and } \rho(r) = \frac{3M}{4\pi b^3}\left(1 + \frac{r^2}{b^2}\right)^{-5/2}.$$

- Using Osipkov and Merritt's method (Osipkov, 1975, and Merritt, 1985) we find the DF is

$$f_Q(Q) = \frac{3Mb^2}{\pi^3\sqrt{2}r_\mathrm{a}^2}\left(\frac{16(r_\mathrm{a}^2 - b^2)}{7}Q^{7/2} + (GM)^2Q^{3/2}\right).$$

where

$$Q = -E - \frac{L^2}{2r_a}.$$

- The PDF for the observables is

$$f_{\boldsymbol{W}_\mathrm{p}}(\boldsymbol{w}_\mathrm{p}; \boldsymbol{a}) = \int_\mathbf{R}\int_\mathbf{R}\int_\mathbf{R} f_Q(Q; \boldsymbol{a})\mathrm{d}z\mathrm{d}v_x, \mathrm{d}v_y.$$

# Recovering the DF parameter

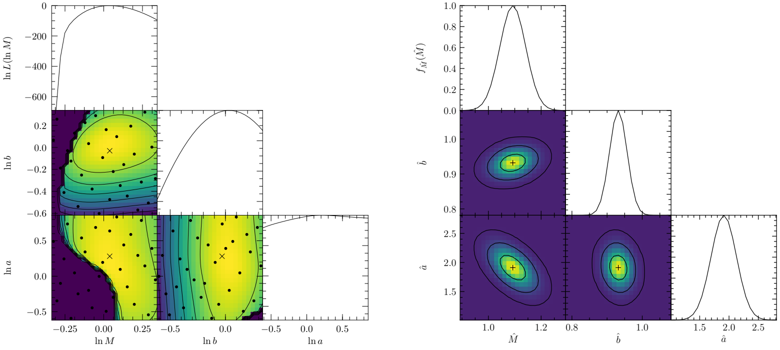



**Figure:** Left: log-marginalized likelihood of the DF parameter. Right: distribution of the maximum-likelihood estimate of the DF parameter (68 %, 95 %, and 95.7 % confidence regions).

# Recovering the parameter

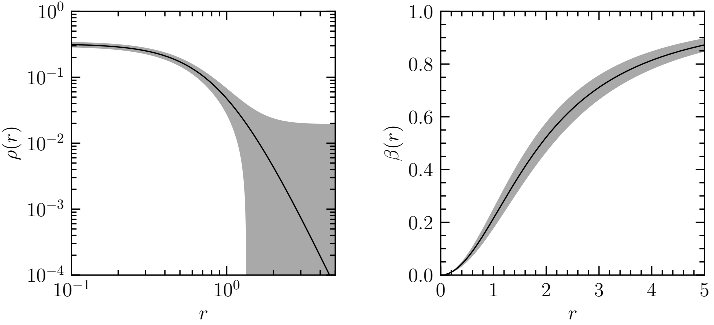



**Figure:** Left: galactic density profile. Right: galactic velocity anisotropy profile, $\beta = 1/(1 + r_a^2/r^2)$.

# Regression

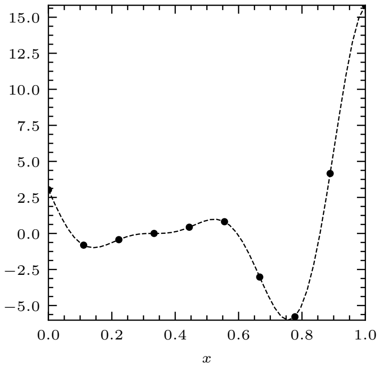



**Figure:** A sample of the Forrester function.

# Regression

**Definition (regression model)**

Treat a function value, $y(\boldsymbol{x})$, as realization of a random variable, $Y(\boldsymbol{x})$. Then assume that

$$Y(\boldsymbol{x}) = r(\boldsymbol{x}) + E(\boldsymbol{x})$$

where $r(\boldsymbol{x}) = \mathrm{E}(Y(\boldsymbol{x}))$, and $E(\boldsymbol{x})$ is a random variable.

# Gaussian random processes

## Definition (random process)

A *random process* is a collection of random variables:

$$Z := (Y(\boldsymbol{x}))_{\boldsymbol{x} \in \boldsymbol{X}}.$$

## Definition (Gaussian random process)

A random process is *Gaussian* if any finite subset of its elements has a multivariate Gaussian distribution. A Gaussian random process is defined by its mean, $r$, and covariance, $k$. We write

$$Z \sim \mathrm{GP}(r, k).$$

# Mean and covariance functions

We assume the zero mean function $r$ such that

$$r(\boldsymbol{x}) = 0.$$

We assume the squared-exponential covariance function, $k$, such that

$$\begin{aligned} k(\boldsymbol{x}, \boldsymbol{x}') &:= \operatorname{cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) \\ &= \sigma^2 \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^{\mathrm{t}} \boldsymbol{M} (\boldsymbol{x} - \boldsymbol{x}')\right) \end{aligned}$$

where $\boldsymbol{M} = \operatorname{diag}(m_{x_1}, m_{x_2}, \ldots, m_{x_D})$.

# Gaussian-process emulation

## Theorem (O'Hagan, 1978)

Let $Z \sim \mathrm{GP}(r, k)$ be a Gaussian random process and let $\boldsymbol{y} = (y(\boldsymbol{x}_1), y(\boldsymbol{x}_2), \dots, y(\boldsymbol{x}_n))$ be drawn from $Z$. Then, for arbitrary $\boldsymbol{x}$, it is the case that

$$Y(\boldsymbol{x}) \mid \boldsymbol{y} \sim N(\hat{r}(\boldsymbol{x}), \hat{\sigma}^2(\boldsymbol{x}))$$

where

$$\begin{aligned} \hat{r}(\boldsymbol{x}) &= r(\boldsymbol{x}) + \boldsymbol{k}^{\mathrm{t}}(\boldsymbol{x})\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{r}) \text{ and} \\ \hat{\sigma}^2(\boldsymbol{x}) &= k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}^{\mathrm{t}}(\boldsymbol{x})\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x}) \end{aligned}$$

and $[\boldsymbol{k}(\boldsymbol{x})]_i = k(\boldsymbol{x}, \boldsymbol{x}_i)$, $[\boldsymbol{K}]_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $[\boldsymbol{r}]_i = r(\boldsymbol{x}_i)$.

# Regression

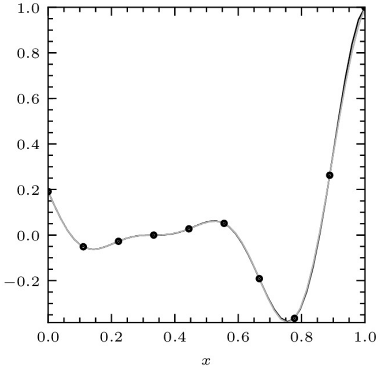



**Figure:** The Forrester function (grey) and its GPE estimate (black).

# GPE literature


- Rasmussen C.E., and C.K.I. Williams, 2006. *Gaussian processes for machine learning*. MIT Press: Cambridge, MA.
- Sacks *et al.*, 1989. Statistical science.
- O'Hagan A., and J.F.C. Kingman, 1978. Journal of the Royal Statistical Society.
- Loeppky J.L., Sack J., and W.J. Welch, 2009. Technometrics.

# Recovering the DF parameter (reprised)

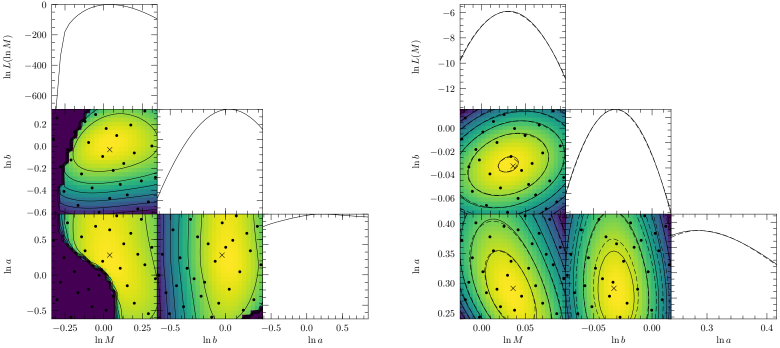



**Figure:** The DF likelihood (dashed lines) and its GPE estimate (solid lines). **Gration A., and M.I. Wilkinson, accepted. MNRAS. https://arxiv.org/abs/1806.06614.**

# Recovering the DF parameter (reprised)

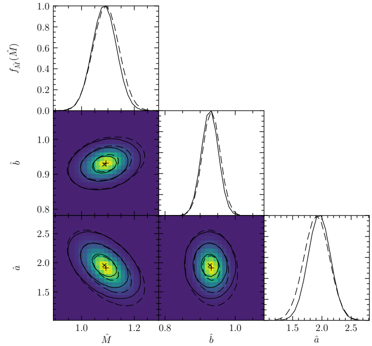



**Figure:** The distribution of the maximum-likelihood estimate of the DF parameter (dashed lines) and its GPE estimate (solid lines) (68 %, 95 %, and 95.7 % confidence regions). **Gration A., and M.I. Wilkinson, accepted. MNRAS. https://arxiv.org/abs/1806.06614.**

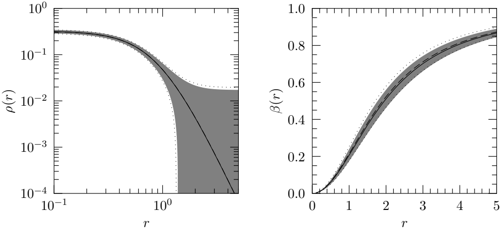



**Figure:** The galactic density profile, velocity anisotropy profile and their 68 % confidence regions (dashed and dotted lines). The GPE estimate of the same (solid lines and grey fills). **Gration A., and M.I. Wilkinson, accepted. MNRAS. https://arxiv.org/abs/1806.06614.**

# Scaling

GPE is suitable for any expensive model.

- Sample size, $N$, is linear in dimension of parameter space, $D$: $N = 10D$ (Loeppky, 2008).
- Fifty-dimensional parameter space is practical.
- Toy model with perfect data is maximally difficult case.

# GPE software

**PyMimic**. Email me for a copy.