

Development and calibration of tumor models

Ernesto A. B. F. Lima
Emanuelle A. Paixão



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES

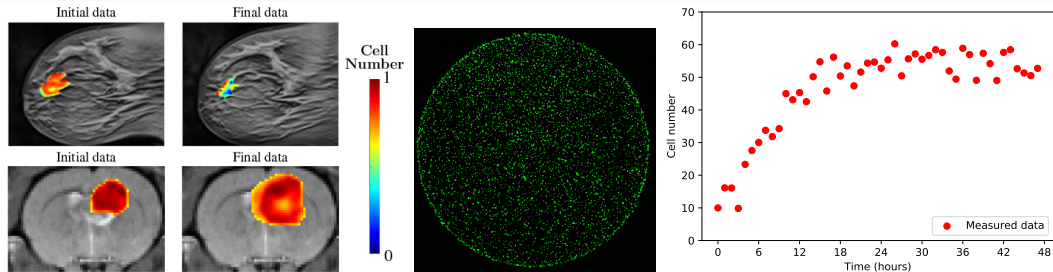


Model Calibration

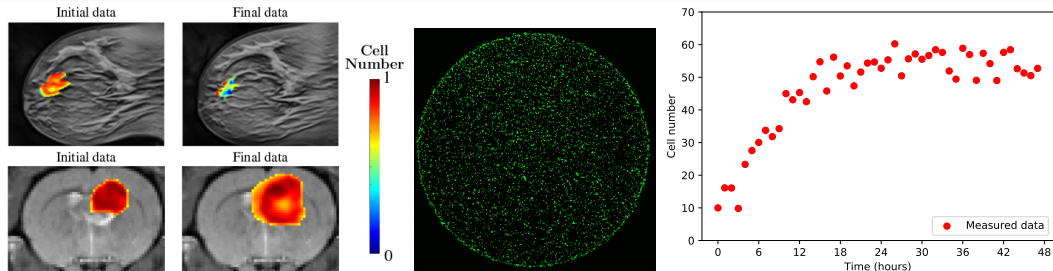
- Calibration Theory:
 - Maximum Likelihood Estimation
 - Bayesian Approach
- EMCEE Python Library¹

¹Foreman-Mackey, Daniel, et al. "emcee v3: A Python ensemble sampling toolkit for affine-invariant MCMC." arXiv preprint arXiv:1911.07688 (2019).

Motivation



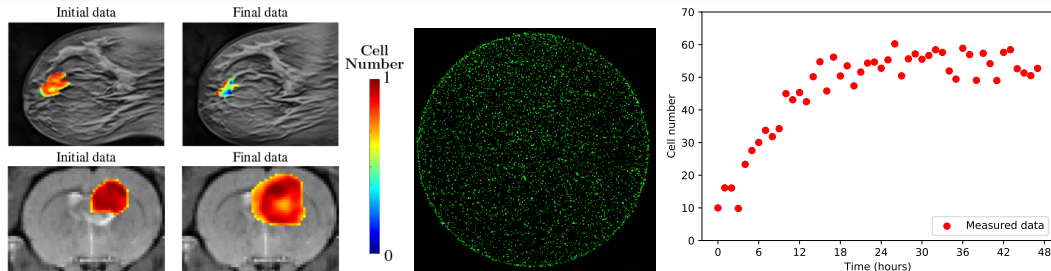
Motivation



$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- N : number of tumor cells;
- r : tumor growth rate;
- K : environment carrying capacity.

Motivation



$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- N : number of tumor cells;
- r : tumor growth rate;
- K : environment carrying capacity.

Model Calibration

It is done to **adjust** the selected model parameters, such as growth and death rates, to obtain the best fit between the **model** predicted responses and the available **data**.

Frequentist and maximum likelihood approaches

- there is an unknown but fixed parameter that represents the event;
- maximum likelihood approach: find the parameter that enables the model to deliver the best characterization of the true target distribution.

²Oden et al., Encyclopedia of Computational Mechanics Second Edition (2017)

Frequentist and maximum likelihood approaches

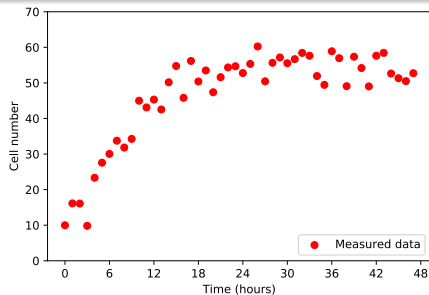
- there is an unknown but fixed parameter that represents the event;
- maximum likelihood approach: find the parameter that enables the model to deliver the best characterization of the true target distribution.

Bayesian approach

- it maps prior beliefs about the parameters to new posterior beliefs in the light of observing the data;
- the estimated parameter is regarded as a random variable, and the true parameter is merely a realization of the random variable.

²Oden et al., Encyclopedia of Computational Mechanics Second Edition (2017)

Definitions



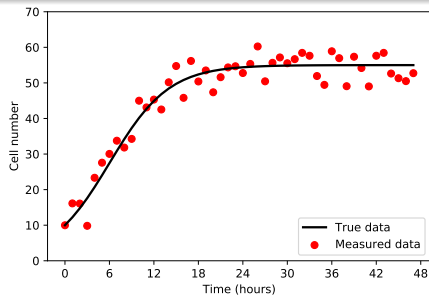
- D : measured data;

Mathematical model

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- θ : vector of model parameters, $\theta = (r, K)$;
- r : tumor growth rate;
- K : environmental carrying capacity;
- $Y(\theta)$: model prediction;

Definitions

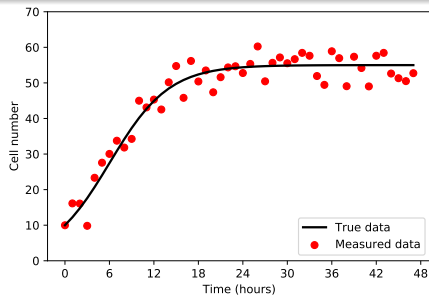


- D : measured data;
- T : true data;

Mathematical model

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- θ : vector of model parameters, $\theta = (r, K)$;
- r : tumor growth rate;
- K : environmental carrying capacity;
- $Y(\theta)$: model prediction;



- D : measured data;
- T : true data;

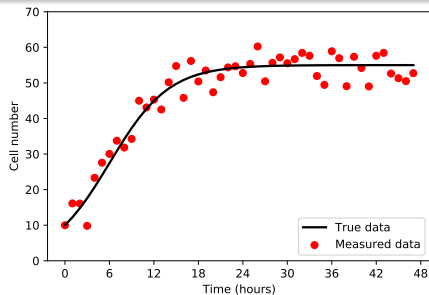
$$T + \epsilon = D,$$

- ϵ : experimental noise;

Mathematical model

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- θ : vector of model parameters, $\theta = (r, K)$;
- r : tumor growth rate;
- K : environmental carrying capacity;
- $Y(\theta)$: model prediction;



- D : measured data;
- T : true data;

$$T + \epsilon = D,$$

- ϵ : experimental noise;

Mathematical model

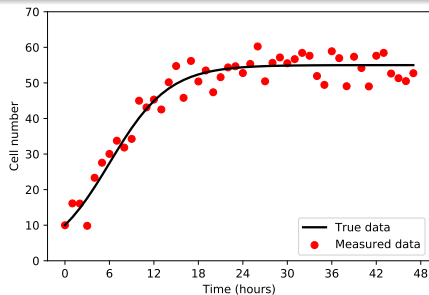
$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- θ : vector of model parameters, $\theta = (r, K)$;
- r : tumor growth rate;
- K : environmental carrying capacity;
- $Y(\theta)$: model prediction;

$$T - Y(\theta) = \gamma(\theta),$$

- γ : model inadequacy.

Definitions



- D : measured data;
 - T : true data;
- $$T + \epsilon = D,$$
- ϵ : experimental noise;

Mathematical model

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

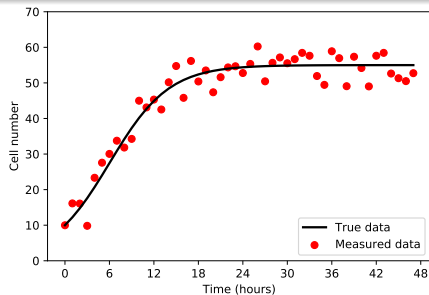
- θ : vector of model parameters, $\theta = (r, K)$;
- r : tumor growth rate;
- K : environmental carrying capacity;
- $Y(\theta)$: model prediction;

$$T - Y(\theta) = \gamma(\theta),$$

$$T - Y^h(\theta) = \gamma^h(\theta),$$

- γ : model inadequacy.

Definitions



- D : measured data;
- T : true data;

$$T + \epsilon = D,$$

- ϵ : experimental noise;

$$D - Y(\theta) = \epsilon + \gamma(\theta),$$

Mathematical model

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

- θ : vector of model parameters, $\theta = (r, K)$;
- r : tumor growth rate;
- K : environmental carrying capacity;
- $Y(\theta)$: model prediction;

$$T - Y(\theta) = \gamma(\theta),$$

$$T - Y^h(\theta) = \gamma^h(\theta),$$

- γ : model inadequacy.

Maximum likelihood estimation

Likelihood

The conditional probability that the data (\mathbf{D}) is observed for a given set of parameters (θ) is the likelihood $\pi(\mathbf{D}|\theta)$.

Likelihood

The conditional probability that the data (\mathbf{D}) is observed for a given set of parameters (θ) is the likelihood $\pi(\mathbf{D}|\theta)$.

$$\mathbf{D} - \mathbf{Y}(\theta) = \epsilon + \gamma(\theta),$$

Assuming:

- 1 the experimental noise is normally distributed ($\epsilon \sim \mathcal{N}(0_{N \times 1}, \sigma_{data}^2 \mathbf{I}_{N \times N})$);
- 2 the model inadequacy is normally distributed ($\gamma \sim \mathcal{N}(0_{N \times 1}, \sigma_{model}^2 \mathbf{I}_{N \times N})$);
- 3 the variance of the total error (σ) is such as $\sigma^2 = \sigma_{data}^2 + \sigma_{model}^2$;
- 4 the data is normally distributed;

Likelihood

The conditional probability that the data (\mathbf{D}) is observed for a given set of parameters ($\boldsymbol{\theta}$) is the likelihood $\pi(\mathbf{D}|\boldsymbol{\theta})$.

$$\mathbf{D} - \mathbf{Y}(\boldsymbol{\theta}) = \boldsymbol{\epsilon} + \boldsymbol{\gamma}(\boldsymbol{\theta}),$$

Assuming:

- 1 the experimental noise is normally distributed ($\boldsymbol{\epsilon} \sim \mathcal{N}(0_{N \times 1}, \sigma_{data}^2 \mathbf{I}_{N \times N})$);
- 2 the model inadequacy is normally distributed ($\boldsymbol{\gamma} \sim \mathcal{N}(0_{N \times 1}, \sigma_{model}^2 \mathbf{I}_{N \times N})$);
- 3 the variance of the total error (σ) is such as $\sigma^2 = \sigma_{data}^2 + \sigma_{model}^2$;
- 4 the data is normally distributed;

$$\pi(\mathbf{D}|\boldsymbol{\theta}) = \prod_{i=1}^{N_t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(D_i - Y_i(\boldsymbol{\theta}))^2}{2\sigma^2}},$$

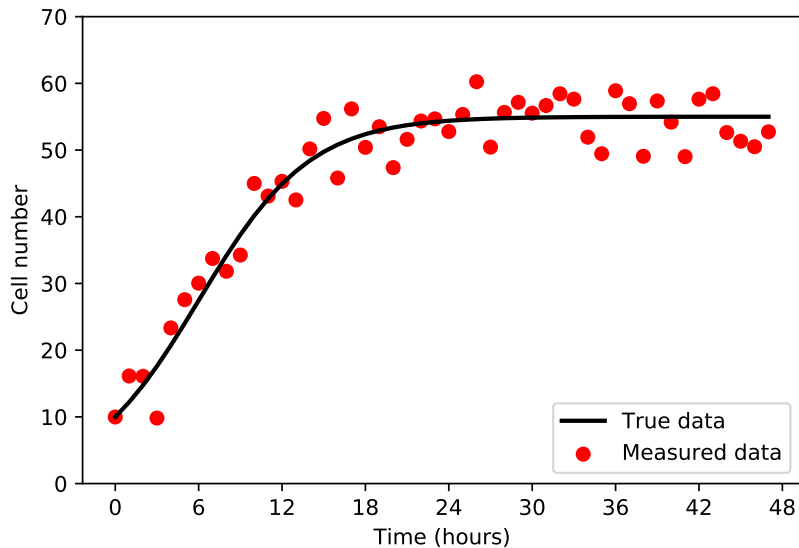
- N_t : the number of data points.

Find $\hat{\boldsymbol{\theta}} \in \Theta$, where Θ is the parameter space, that maximize the likelihood.

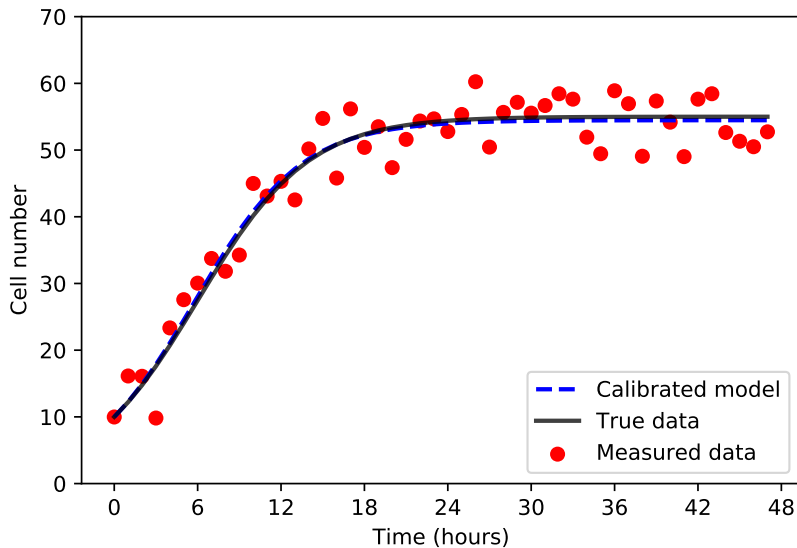
It is customary to work with the more manageable log-likelihood function.

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} [\log \pi(\mathbf{D}|\boldsymbol{\theta})]; \\ &= \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \left[-\frac{1}{2} \sum_{i=1}^{N_t} \left(\log(2\pi) + \log(\sigma^2) + \frac{(D_i - Y_i(\boldsymbol{\theta}))^2}{\sigma^2} \right) \right].\end{aligned}$$

Maximum likelihood estimation



Maximum likelihood estimation



$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right),$$

$$r = 0.25$$

$$K = 59.14$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$

$$P(A, B) = P(B, A);$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$

$$P(A, B) = P(B, A);$$

$$P(B, A) = P(B|A)P(A);$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$

$$P(A, B) = P(B, A);$$

$$P(B, A) = P(B|A)P(A);$$

$$P(A|B)P(B) = P(B|A)P(A);$$

Bayesian approach

Given events A and B:

$$P(A, B) = P(A|B)P(B);$$

$$P(A, B) = P(B, A);$$

$$P(B, A) = P(B|A)P(A);$$

$$P(A|B)P(B) = P(B|A)P(A);$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)};$$

Bayesian approach

Given events A and B :

$$P(A, B) = P(A|B)P(B);$$

$$P(A, B) = P(B, A);$$

$$P(B, A) = P(B|A)P(A);$$

$$P(A|B)P(B) = P(B|A)P(A);$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)};$$

Converting to probability densities π , if A represents the parameter θ of a model, and B the observational data \mathbf{D} :

$$\underbrace{\pi(\theta|\mathbf{D})}_{\text{posterior}} = \frac{\overbrace{\pi(\mathbf{D}|\theta)}^{\text{likelihood}} \overbrace{\pi(\theta)}^{\text{prior}}}{\underbrace{\pi(\mathbf{D})}_{\text{evidence}}};$$

$$\pi(\mathbf{D}) = \int_{\Theta} \pi(\mathbf{D}|\theta)\pi(\theta) d\theta$$

Bayesian approach

Given events A and B:

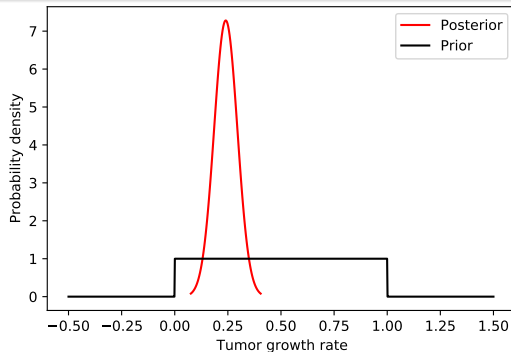
$$P(A, B) = P(A|B)P(B);$$

$$P(A, B) = P(B, A);$$

$$P(B, A) = P(B|A)P(A);$$

$$P(A|B)P(B) = P(B|A)P(A);$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)};$$

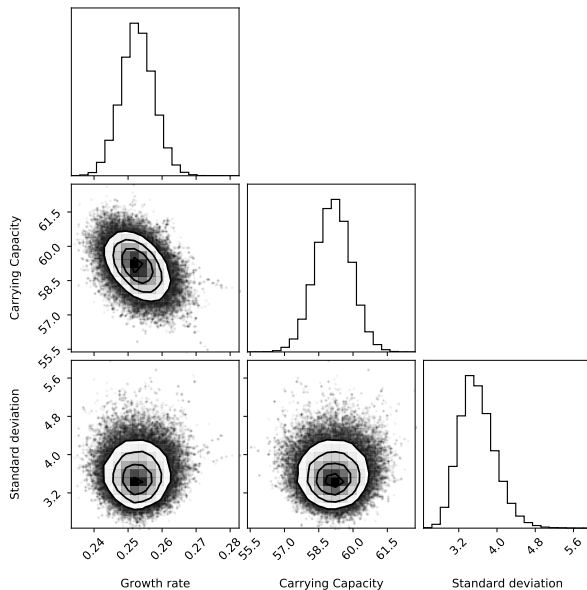


Converting to probability densities π , if A represents the parameter θ of a model, and B the observational data \mathbf{D} :

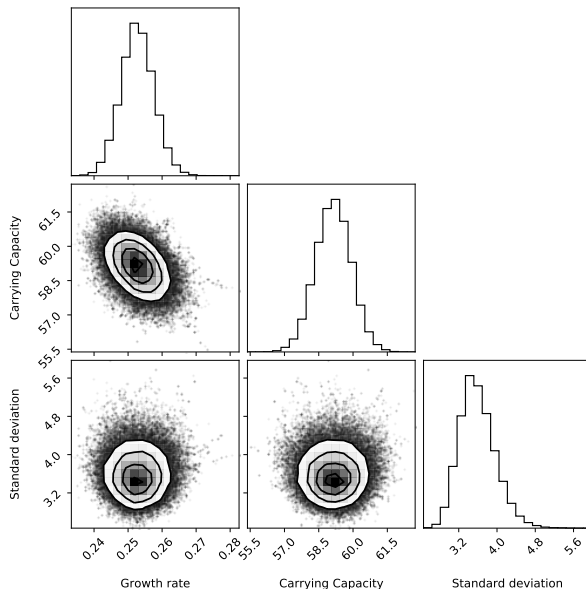
$$\underbrace{\pi(\theta|\mathbf{D})}_{\text{posterior}} = \frac{\overbrace{\pi(\mathbf{D}|\theta)}^{\text{likelihood}} \overbrace{\pi(\theta)}^{\text{prior}}}{\underbrace{\pi(\mathbf{D})}_{\text{evidence}}};$$

$$\pi(\mathbf{D}) = \int_{\Theta} \pi(\mathbf{D}|\theta)\pi(\theta) d\theta$$

Bayesian approach

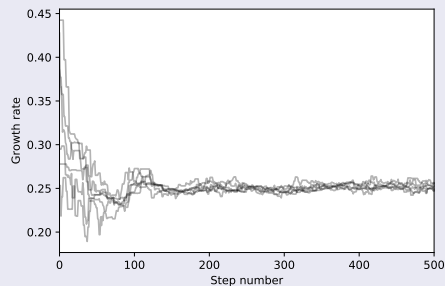


Bayesian approach

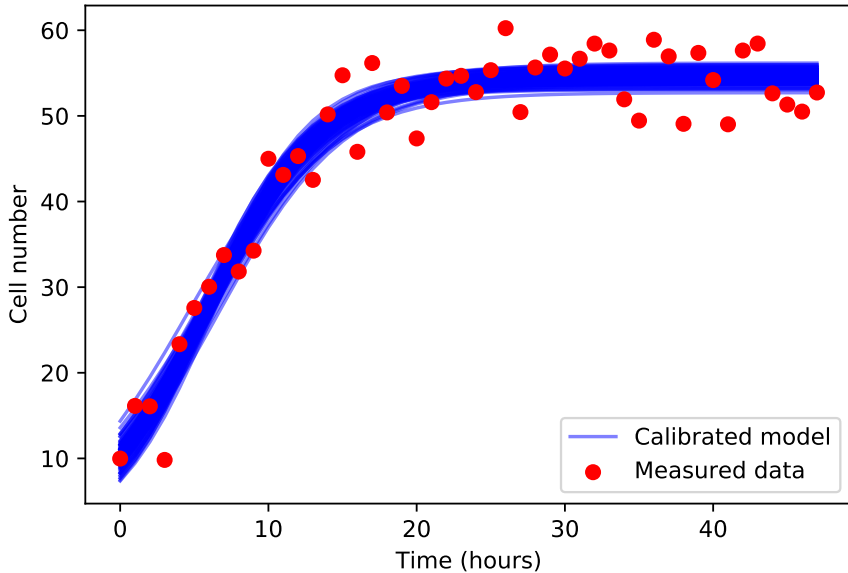


Markov Chain Monte Carlo:

Markov Chain Monte Carlo methods draw samples where the next sample is dependent on the existing sample, called a Markov Chain.



Bayesian approach



Jupyter Notebook - Binder and Discord

<https://github.com/ernesto-lima/veraolncc2021>

The screenshot shows a Discord server interface for the channel #mini-course-1. The server is named EMMCCT 2021. The channel name is #mini-course-1, with a description: "MC01-CT: Development and Calibration of Tumor Models (Desenvolvimento e Calibração de Modelos Tumorais)".

Left Sidebar (Server Channels):

- ADMIN
 - admin
 - admin-messages
 - Admin Voice Channel
- PANELIST
 - panelist-messages
 - Panelist Voice Channel
- EMMCCT 2021
 - helpdesk
 - announcements
 - event-agenda
 - posters
 - #mini-course-1** (selected)
 - mini-course-2
 - random-talk

Main Channel Content:

- A large "#" symbol.
- Bem-vindo(a) a #mini-course-1!**
- Este é o começo do canal #mini-course-1. MC01-CT: Development and Calibration of Tumor Models (Desenvolvimento e Calibração de Modelos Tumorais).
- [Editar canal](#)
- A message from **Anna Claudia Resende** (22/01/2021):
 - Professores:** Ernesto Lima (UT at Austin) e Emanuele Arantes Paixão (LNCC).
 - Horários:** De quarta (27/01) a sexta (29/01) das 11:00h às 12:30h.
 - Os principais objetivos deste minicurso são:**
 - Familiarizar os participantes com a motivação biológica para o uso de diversos modelos de crescimento tumoral;
 - Apresentar métodos de resolução de modelos de equações diferenciais ordinárias (EDO's), utilizando Python;
 - Calibrar os modelos desenvolvidos via métodos Bayesianos.
 - Ementa:**
 - Desenvolvimento da parte teórica sobre modelagem em câncer (27/01): uma introdução sobre modelagem no contexto do câncer, abordando modelos discretos, contínuos e híbridos. Exemplificação e interpretação de diferentes tipos de funções para a descrição do crescimento tumoral, mortalidade, efeito Allee e terapias, entre outros fenômenos biológicos;
 - Introdução à linguagem de programação Python (28/01): resolução de EDO's e diferentes tipos de função, trabalhadas durante o primeiro dia de curso, em Python;
 - Desenvolvimento da parte teórica e prática sobre calibração (29/01): abordagem de conceitos relacionados à calibração, de forma mais específica à calibração Bayesiana. Os conteúdos teóricos e práticos serão abordados conjuntamente, por meio da calibração de um modelo de EDO, utilizando dados gerados. Nesta atividade, será trabalhada a resolução da EDO e construção de gráficos com os resultados obtidos, usando a biblioteca emcee para calibrar o modelo.

Right Sidebar (User List):

- HELPER--2
 - Anna Claudia Resende
 - Emanuelle Paixão
- ORGANIZER--2
 - Luciana Barros
 - Regina Almeida
- PANELIST--1
 - Heber Rocha
- DISPONIVEL--7
 - AlejandroHerrera
 - eye5
 - FelipeCandian
 - LeoSouza
 - Oscar Antezana
 - Valerio
 - VzMedivh
- OFFLINE--52
 - _lucasDC_
 - AlejandroHerrera
 - Alfredo Scari
 - Allan Costa