Use of collection of models and model selection

Stéphanie Portet

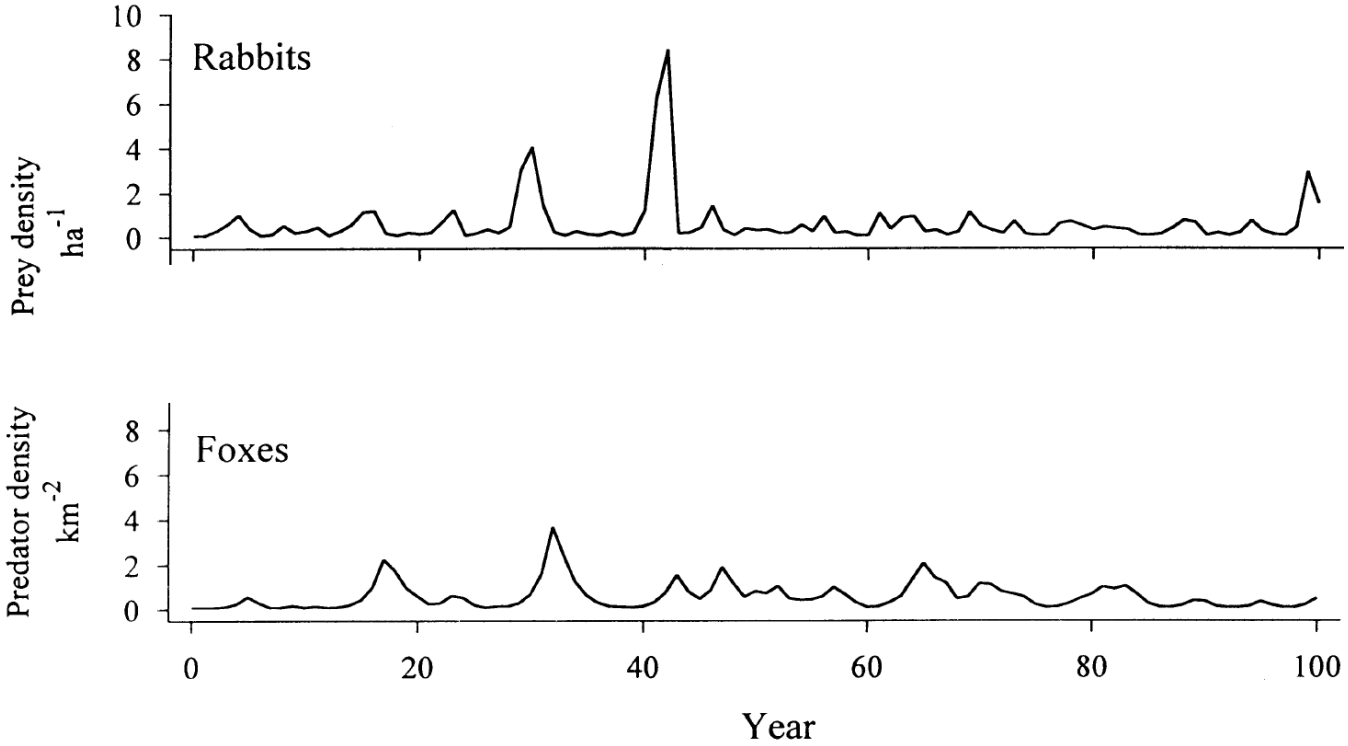
Department of Mathematics
University of Manitoba
Winnipeg, MB, Canada
Stephanie.Portet@umanitoba.ca

March 2023

Outline

1. Combining data and mathematical models - how to get more
2. Example: Interactions between mTOR and NMT1 in breast cancer cells
3. Conclusion

Example: Rabbits and foxes in Australia



$$\frac{dR}{dt} = aR - bRF \qquad \frac{dR}{dt} = a\left(1 - \frac{R}{k}\right)R - bRF \qquad \frac{dR}{dt} = aR - \frac{bRF}{1 + fR}$$

$$\frac{dF}{dt} = -cF + eRF \qquad \frac{dF}{dt} = -cF + eRF \qquad \frac{dF}{dt} = -cF + \frac{eRF}{1 + fR}$$

Which model of the collection would represent the best the reality given the data we have recorded?

Why consider a collection of models

Lack of knowledge

- considering different assumptions for a given mechanism
- different translations of the same assumption

Design a collection of models

- to mimic the well known positive and negative control experimental protocol used in experimental labs
- to allow the identification of a most plausible scenario for the defined problem

How to identify the best model of the collection? $\Rightarrow$ Model selection

- Information Theory Criteria
- Statistical Tests (nested models)

Nested or non-nested models

$$\frac{dR}{dt} = aR - bRF \qquad \frac{dR}{dt} = a\left(1 - \frac{R}{k}\right)R - bRF \qquad \frac{dR}{dt} = aR - \frac{bRF}{1+fR}$$
$$\frac{dF}{dt} = -cF + eRF \qquad \frac{dF}{dt} = -cF + eRF \qquad \frac{dF}{dt} = -cF + \frac{eRF}{1+fR}$$

Nested models: the three models are particular cases of the full model

$$\frac{dR}{dt} = aR - vR^2 - \frac{bRF}{1+fR}$$
$$\frac{dF}{dt} = -cF + \frac{eRF}{1+fR}$$

$v = f = 0 \Rightarrow$ Lotka-Volterra model

$f = 0 \Rightarrow$ Model with logistic dynamics for preys

$v = 0 \Rightarrow$ Model with saturating rate

Model selection

Consider a collection of R models.

Which model of the collection would represent the best the reality given the data we have recorded?

In 1973, Akaike found a relationship between the maximum likelihood (statistical analysis) and Kullback-Leibler divergence (information theory).

Akaike (1973) In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, pp 267–281

Kullback-Leibler (KL) divergence

To measure the difference between two probability distribution functions $f(x)$ and $g(x)$ over the same variable x

The **KL divergence of** $g(x)$ **from** $f(x)$ is

$$I(f,g) = \int_X f(x) \ln\left(\frac{f(x)}{g(x)}\right) dx$$

if $f(x)$ and $g(x)$ two pdfs of a continuous random variable

Properties

- not symmetric $I(f,g) \neq I(g,f)$
- $I(f,g) \geq 0$
- $I(f,g) = 0$ iff $f = g$

Kullback and Leibler (1951) The Annals of Mathematical Statistics.

Kullback-Leibler divergence

KL divergence = measure of the information lost when approximating the full reality $f(x)$ by a model $g(x|\theta)$

$$I(f,g) = \int f(x) \ln\left(\frac{f(x)}{g(x|\theta)}\right) dx$$

where the model $g(x|\theta)$ depends on parameter θ

Problems:

- the reality/truth f is unknown
- θ must be estimated from data y (generated from f)

Kullback and Leibler (1951) The Annals of Mathematical Statistics. 22: 79-86.

Measurement of the information lost when approximating the full reality $f(x)$ by a model $g(x|\theta)$

$$
\begin{aligned}
I(f,g) &= \int f(x) \ln\left(\frac{f(x)}{g(x|\theta)}\right) dx \\
&= \int f(x) \ln(f(x)) dx - \int f(x) \ln(g(x|\theta)) dx \\
&= E_f[\ln(f(x))] - E_f[\ln(g(x|\theta))] \\
&= C - E_f[\ln(g(x|\theta))]
\end{aligned}
$$

where $E_f[\ln(f(x))] = C$ depends only on the unknown true distribution f and is unknown

$$I(f,g) - C = -E_f[\ln(g(x|\theta))]$$

Relative KL divergence between f and $g = -E_f[\ln(g(x|\theta))]$

Comparison of models

Consider two models g_1 and g_2.

If $I(f, g_1) < I(f, g_2)$ then the model g_1 is better than model g_2

$$\begin{aligned} I(f, g_1) &< I(f, g_2) \\ I(f, g_1) - C &< I(f, g_2) - C \\ -E_f[\ln(g_1(x|\theta))] &< -E_f[\ln(g_2(x|\theta))] \end{aligned}$$

$$\Rightarrow I(f, g_2) - I(f, g_1) = -E_f[\ln(g_2(x|\theta))] + E_f[\ln(g_1(x|\theta))]$$

Without knowing C, we know how much better g_1 is than g_2

Using relative KL divergence between f and g_i, we can compare the models g_i

- $\hat{\theta}(y)$ estimator of θ = random variable
- $I(f, g(.|\hat{\theta}(y)))$ = random variable
- $E_y\left[I(f, g(.|\hat{\theta}(y)))\right] = C - E_y\left[E_x\left[\ln(g(x|\hat{\theta}(y)))\right]\right]$
  where x and y are independent random samples from the same distribution and both statistical expectations are taken with respect to truth f

Aim: minimize the estimated expected KL divergence over a set of models considered $\Leftrightarrow$ maximize the estimated expected relative KL divergence

Model selection criterion

$$\max_{g \in G} E_y\left[E_x\left[\ln(g(x|\hat{\theta}(y)))\right]\right]$$

where G is the collection of models in terms of probability density functions

Akaike Information Criterion

In 1973, Akaike found

- an asymptotically (for large sample) unbiased estimator of the expected relative Kullback-Leibler divergence that is

$$\ln \mathcal{L}(\hat{\theta}_{MLE}|y) - K$$

where $\mathcal{L}$ is the likelihood function, $\hat{\theta}_{MLE}$ is the maximum likelihood estimate of θ and K is the number of estimated parameters

Akaike Information Criterion for each model considered with the same data set is defined as

$$AIC = -2 \ln \left(\mathcal{L}(\hat{\theta}_{MLE}|y) \right) + 2K$$

Best model = the one with minimum AIC value

Akaike (1973) In: Petrov BN, Csaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, pp 267–281

Akaike Information Criterion

Consider a collection of R models: Which model of the collection would represent the best the reality given the data we have recorded?

Compute for each model the information criterion:

If $K < (N/40)$, use AIC

$$AIC = -2 \ln \left(\mathcal{L}(\hat{\theta}_{MLE}|y) \right) + 2K$$

If $K > (N/40)$, use corrected AIC (AICc)

$$AICc = -2 \ln \left(\mathcal{L}(\hat{\theta}_{MLE}|y) \right) + 2K + \frac{2K(K+1)}{N-K-1} = AIC + \frac{2KN}{N-K-1}$$

where K is the number of estimated parameters and N is the number of observations (sample size)

As $N \to \infty$, $AICc \to AIC$

Sugiura (1978) Communications in Statistics - Theory and Methods. 7:13. Hurvich and Tsai (1989) Biometrika 76:297.

When the **measurement errors are independent and identically normally distributed with the same variance**

$$-\ln \mathcal{L}(\hat{\theta}_{MLE}|y) = \frac{N}{2}\ln(2\pi) + \frac{N}{2}\ln\left(\frac{F_{ls}(\hat{p}_{MLE})}{N}\right) + \frac{N}{2}$$

where $\hat{p}_{MLE} = \hat{p}_{LSE}$ $(N = n_e n_o^e n_i^{e,o})$

When **the data used to compare all the models are the same**, AIC can be computed as follows

$$AIC = N\ln\left(\frac{F_{ls}(\hat{p}_{LSE})}{N}\right) + 2K$$

where K is the number of estimated parameters (number of estimated mathematical model parameters + 1), N is the number of observations

Compute AIC_i of each model i with $i \in \{1, \ldots, R\}$
Best model = the one with minimum AIC value

AIC differences

As only the estimates of the expected relative K-L divergences between f and $g_i(x|\theta)$ are known with the information criteria, it is convenient to scale them with respect to the minimum AIC value among all models.

AIC differences: estimate of information loss when using model i rather than the estimated best model

$$\Delta_i = AIC_i - \min_i AIC_i$$

$\min_i AIC_i$ = AIC of the best model in the collection

Interpretation = the larger the Δ_i, the less plausible is model i

Akaike (1974) IEEE Trans. Automatic Control. 19:716. Burnham and Anderson (2002) Model selection and multimodel inference: a practical information-theoretic approach. Second Edition. Springer.

Interpretation = the larger the Δ_i, the less plausible is model i

Some guidelines for nested models:

- $\Delta_i \in \{1, 2\}$ model i has substantial support and should be considered
- $\Delta_i \in \{4, \ldots, 7\}$ model i has less support
- $\Delta_i > 10$ model i has no support, can be omitted

Might be different for non-nested models or for a very large number of models

Burnham and Anderson (2002) Model selection and multimodel inference: a practical information-theoretic approach. Second Edition. Springer.

Akaike weights

For an easier interpretation, rescaling of Δ_i

Likelihood of model i given the data $\propto \exp\left(-\frac{\Delta_i}{2}\right)$

Akaike weight or "weight of evidence" of model i for being the best model of the collection given the data

$$w_i = \frac{exp(-\Delta_i/2)}{\sum_{r=1}^{R} exp(-\Delta_r/2)}$$

w_i = probability that model i is the best (approximating) model given the experimental data and the collection of models considered

Interpretation

- The smaller the weight w_i, the less plausible is model i
- Consider a single best model i if $w_i > 0.9$

Uses of Akaike weights

Evidence ratio of model i versus model j = Strength of evidence in favour of model i over model j

$$\frac{w_i}{w_j}$$

Confidence set of models:
Sum the Akaike weights from largest to smallest until the sum is ≥ 0.95 $\Rightarrow$ the corresponding subset of models is the 95% confidence set on the best model

Relative importance of a process:
Sum the Akaike weights over all models in which the process of interest appears = measure of the relative importance of the process of interest

Model selection with Akaike Information Criterion

- Best model = the one with the lowest AIC
- Best model within the collection of model considered given the experimental data $\neq$ "true model"
- No meaning in the actual values of AIC
- Ranking of candidate models
- Selection of a model with the least number of parameters that best-fits experimental data
- Specific to a given set of data (cannot be used to compare models on different data sets)
- Valid to compare nested or non-nested models
- Not a test!!

Portet (2020) Inf. Dis. Model. 5.

Strategy as using data and mathematical modelling

- Systematic modelling of possible scenarios based on biological hypotheses and first principles to design a collection of models
- Calibration of each model using **the same data**
- Compute AIC and Akaike weights for each model $\Rightarrow$ rank models and identify the best model or the 95% confidence set of models
- Partition the collection of models in subsets of models based on their underlying hypotheses and using Akaike weights, evaluate the importance of different processes

Portet *et al.* (2015) PLOS One. Jacquier *et al.* (2018) Scientific Report. Lee *et al.* (2021) AIMS Mathematics. Portet (2020) Infectious Disease Modelling.

Outline

1. Combining data and mathematical models - how to get more
2. Example: Interactions between mTOR and NMT1 in breast cancer cells
3. Conclusion

Motivation

mTOR regulates cell growth, proliferation, survival and migration

- kinase (phosphorylated other molecules)
- 2 forms of mTOR: inactive (mTOR) and active (pmTOR = phosphorylated-mTOR)

Cancer cells exploit mTOR to enhance their capacity to growth

Strong expression of NMT1 has been reported in malignant breast tissues compared with normal breast cells

- enzyme
- 2 forms of NMT1: inactive (pNMT1) and active (NMT1)

$\Rightarrow$ Interactions between NMT1 and mTOR

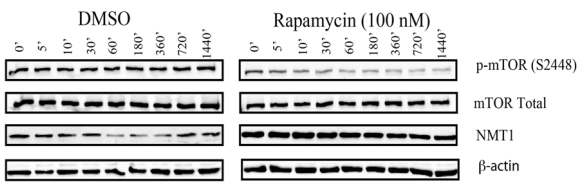
Work in collaboration with A. Shrivastav (University of Winnipeg, Manitoba, Canada)

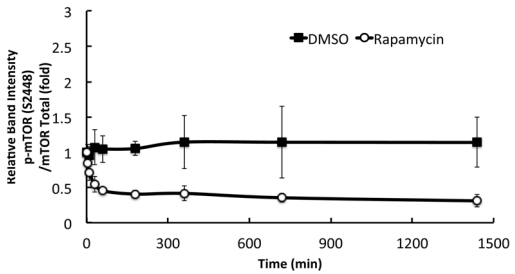
Experimental data

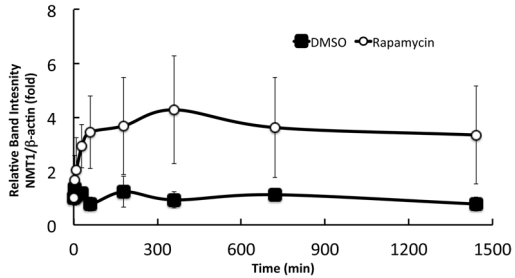
Four datasets

- Total mTOR = mTOR + pmTOR
- pmTOR
- Total NMT1 = NMT1 + pNMT1

at nine time points under control (DMSO) and perturbed (Rapamycin) conditions







- Rapamycin treatment decreases the phosphorylation of mTOR (pmTOR) and augments total NMT1 levels over time
- No significant change in the total mTOR levels under both experimental conditions

Rapamycin = drug used in cancer, targets mTOR and prevents its activation (prevents mTOR phosphorylation)

To investigate the regulation of NMT1 by mTOR

Combine experimental data and mathematical modelling

- core assumption: NMT1 phosphorylation is regulated by pmTOR
- system under two conditions
  - unperturbed system = without rapamycin (control)
  - perturbed system = with rapamycin (inhibition of mTOR)

Collection of models to test alternative hypotheses

- Does the regulation of endogenous levels of mTOR components impact the dynamics?
  - synthesis and degradation **vs** constant Total mTOR
- Does NMT1 have a negative feedback effect on mTOR?
  - feedback **vs** no feedback
- Is the effect of rapamycin on mTOR reversible or irreversible?
  - reversible **vs** irreversible binding

Jacquier, M. *et al.* (2018) Scientific Reports 8:12969.

Model framework

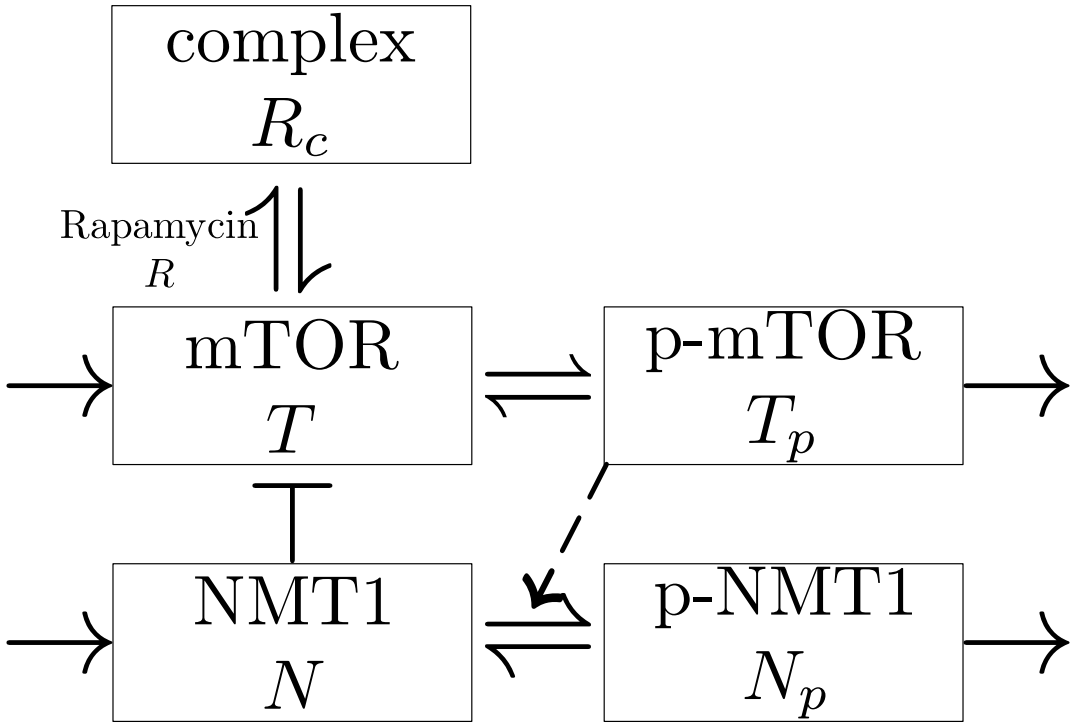
$$\frac{\mathrm{d}T}{\mathrm{d}t} = \overbrace{\frac{-\alpha_T T}{K_T + T}}^{\text{phosphorylation}} + \overbrace{\frac{\alpha_{T_p} T_p}{K_{T_p} + T_p}}^{\text{dephosphorylation}} + \overbrace{\Pi_T}^{\text{synthesis}} + \overbrace{f(T,N)}^{\text{feedback}} + \overbrace{g(T,R_c)}^{\text{rapamycin effect}},$$

$$\frac{\mathrm{d}T_p}{\mathrm{d}t} = \overbrace{\frac{\alpha_T T}{K_T + T}}^{\text{phosphorylation}} - \overbrace{\frac{\alpha_{T_p} T_p}{K_{T_p} + T_p}}^{\text{dephosphorylation}} - \overbrace{\delta_{T_p} T_p}^{\text{degradation}},$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \overbrace{-\frac{\alpha_N T_p N}{K_N + N}}^{\text{phosphorylation}} + \overbrace{\Pi_N}^{\text{synthesis}},$$

$$\frac{\mathrm{d}N_p}{\mathrm{d}t} = \overbrace{\frac{\alpha_N T_p N}{K_N + N}}^{\text{phosphorylation}} - \overbrace{\delta_{N_p} N_p}^{\text{degradation}},$$

$$\frac{\mathrm{d}R_c}{\mathrm{d}t} = \overbrace{h(T,R_c)}^{\text{rapamycin effect}}.$$



Control (DMSO) $\Rightarrow g(T, R_C) = h(T, R_C) = 0$

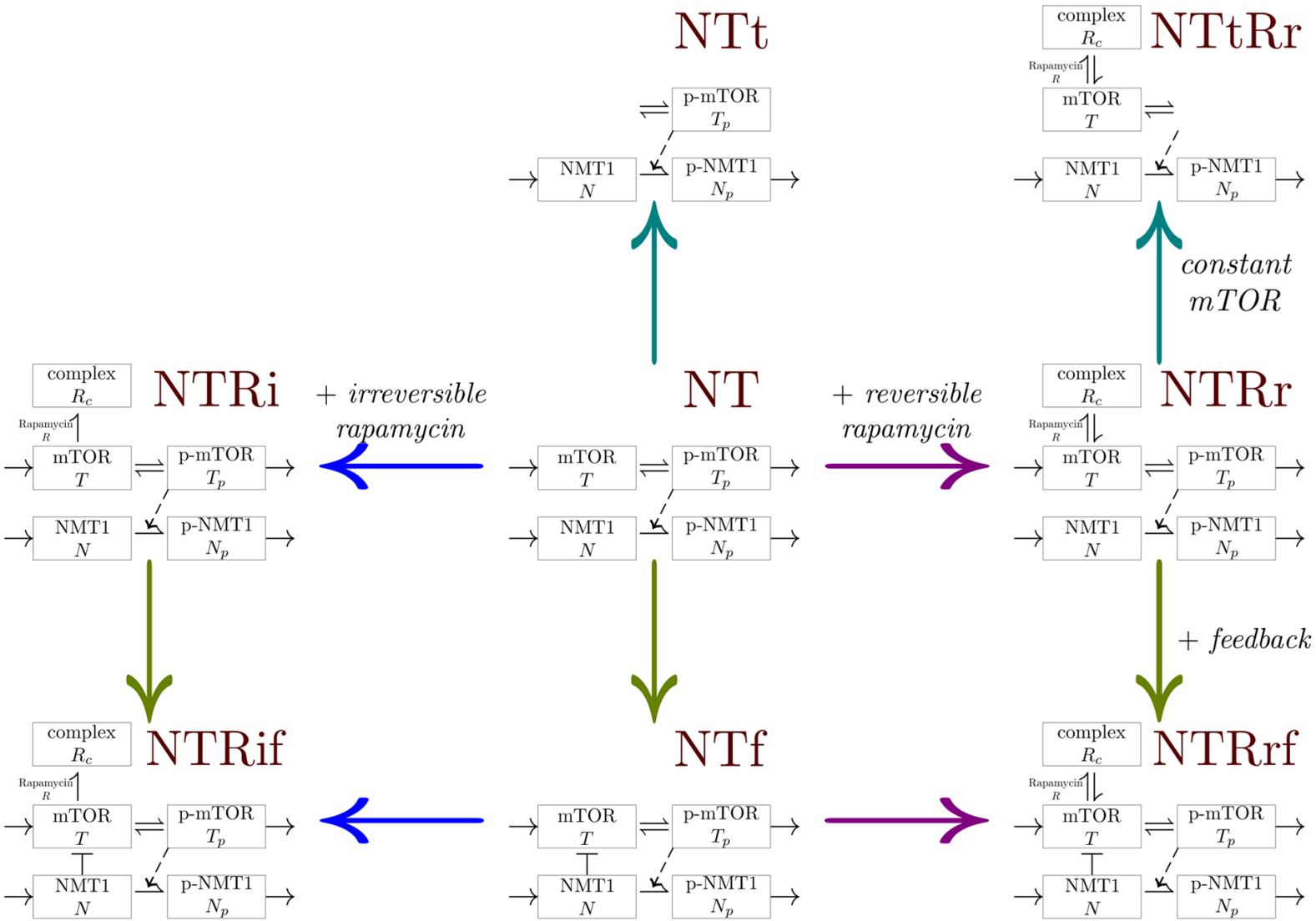
Negative feedback $\Rightarrow f(T, N) = -\beta TN$

Collection of eight models

Perturbed system

Unperturbed system

Perturbed system



Jacquier, M. *et al.* (2018) Scientific Reports 8:12969.

Model Calibration

For each of the four datasets

- NT, NTt and NTf models are calibrated to data with DMSO
- NTRr, NTtRr, NTRi, NTRrf and NTRif models are calibrated to data with rapamycin

Model	p	rapamycin	feedback	mTOR
NT	10	no	no	s/d
NTt	9	no	no	constant
NTf	11	no	yes	s/d
NTRr	12	reversible	no	s/d
NTtRr	11	reversible	no	constant
NTRi	12	irreversible	no	s/d
NTRrf	13	reversible	yes	s/d
NTRif	13	irreversible	yes	s/d

- For each model i, minimize

$$\mathrm{RSS}_i = \sum_{j=1}^{m} \overbrace{\left(T_p^{exp}(t_j) - T_p^i(t_j)\right)^2}^{pmTOR} + \overbrace{\left(T_{total}^{exp}(t_j) - T_{total}^i(t_j)\right)^2}^{pmTOR+mTOR} + \overbrace{\left(N_{total}^{exp}(t_j) - N_{total}^i(t_j)\right)^2}^{pNMT1+NMT1}$$

and $T_{total}^i = T^i + T_p^i$ and $N_{total}^i = N^i + N_p^i$ for model i

Results of calibration

For the 4 datasets

- Unperturbed system: All models (NT, NTt and NTf) reproduce the trends in the proportion of pmTOR and total NMT1 observed for all experimental datasets without rapamycin
- Perturbed system: All models (NTRr, NTtRr, NTRi, NTRrf and NTRif) reproduce the trends observed in experimental data with rapamycin, in particular a decrease in the proportion of p-mTOR and an increase in total NMT1

Model responses support the regulation of NMT1 by pmTOR in the presence or absence of rapamycin

Model selection

Use of the AIC corrected for small sample sizes instead of the AIC

$$\mathrm{AICc}_i = \mathrm{AIC}_i + \frac{2K_i(K_i+1)}{N-K_i-1} = N \ln\left(\frac{\mathrm{RSS}_i}{N}\right) + 2K_i \frac{N}{N-K_i-1}$$

as the number of data points is small ($N = 3m = 27$ for each dataset) in comparison to the number of parameters $K_i = p_i + 1 =$ number of estimated parameters for model i including the estimation of RSS_i/N

For each model i, we have 4 values of AICc_i as we have 4 different datasets

Sugiura (1978) Communications in Statistics - Theory and Methods. Hurvich and Tsai (1989) Biometrika.

Results (1/2)

Model	rapamycin	feedback	mTOR
NT	no	no	s/d
NTt	no	no	constant
NTf	no	yes	s/d
NTRr	revers.	no	s/d
NTtRr	revers.	no	constant
NTRi	irrevers.	no	s/d
NTRrf	revers.	yes	s/d
NTRif	irrevers.	yes	s/d

Models		Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	k_i	$AICc_i$	w_i	$AICc_i$	w_i	$AICc_i$	w_i	$AICc_i$	w_i
(a) Without rapamycin									
NT	11	−6.9	0.062	−101.9	0.033	−67.2	0.045	−109.2	0.379
NTt	10	−12.4	**0.936**	−108.6	**0.967**	−73.5	**0.957**	−110.2	**0.609**
NTf	12	−0.7	0.003	−92.9	10^{-4}	−61.2	0.002	−102.4	0.012
(b) With rapamycin									
NTRr	13	−2.3	0.393	−107.1	10^{-4}	−75.9	0.025	−58.8	0.407
NTtRr	12	−1.4	0.248	−124.9	**0.997**	−83.1	**0.929**	−58.2	0.305
NTRi	13	−2.0	0.351	−113.5	0.003	−76.7	0.039	−58.1	0.282
NTRrf	14	7.0	0.004	−91.6	10^{-8}	−73.2	0.007	−48.0	0.002
NTRif	14	6.6	0.005	−102.4	10^{-5}	−66.6	10^{-4}	−49.4	0.004

Unperturbed system: NTt is the best model for datasets 1, 2 and 3 (non-conclusive for dataset 4)

Perturbed system: NTtRr (analogue of NTt with reversible binding for rapamycin) is the best model for the datasets 2 and 3 (non conclusive for datasets 1 and 4)

Results (2/2)

Model	rapamycin	feedback	mTOR
NT	no	no	s/d
NTt	no	no	constant
NTf	no	yes	s/d
NTRr	revers.	no	s/d
NTtRr	revers.	no	constant
NTRi	irrevers.	no	s/d
NTRrf	revers.	yes	s/d
NTRif	irrevers.	yes	s/d

Weights corresponding to each assumption are obtained by summing the weights of the models verifying the assumption

Models		Dataset 1		Dataset 2		Dataset 3		Dataset 4	
	k_i	$AICc_i$	w_i	$AICc_i$	w_i	$AICc_i$	w_i	$AICc_i$	w_i
(a) Without rapamycin									
NT	11	−6.9	0.062	−101.9	0.033	−67.2	0.045	−109.2	0.379
NTt	10	−12.4	**0.936**	−108.6	**0.967**	−73.5	**0.957**	−110.2	**0.609**
NTf	12	−0.7	0.003	−92.9	10^{-4}	−61.2	0.002	−102.4	0.012
Assumptions									
Feedback			0.003		0.0004		0.002		0.01
No feedback			**0.997**		**0.9996**		**0.998**		**0.99**
(b) With rapamycin									
NTRr	13	−2.3	0.393	−107.1	10^{-4}	−75.9	0.025	−58.8	0.407
NTtRr	12	−1.4	0.248	−124.9	**0.997**	−83.1	**0.929**	−58.2	0.305
NTRi	13	−2.0	0.351	−113.5	0.003	−76.7	0.039	−58.1	0.282
NTRrf	14	7.0	0.004	−91.6	10^{-8}	−73.2	0.007	−48.0	0.002
NTRif	14	6.6	0.005	−102.4	10^{-5}	−66.6	10^{-4}	−49.4	0.004
Assumptions									
Reversible			**0.64**		**0.9967**		**0.961**		**0.714**
Irreversible			0.36		0.0033		0.039		0.286
Feedback			0.008		10^{-5}		0.007		0.006
No feedback			**0.992**		**0.99999**		**0.993**		**0.994**

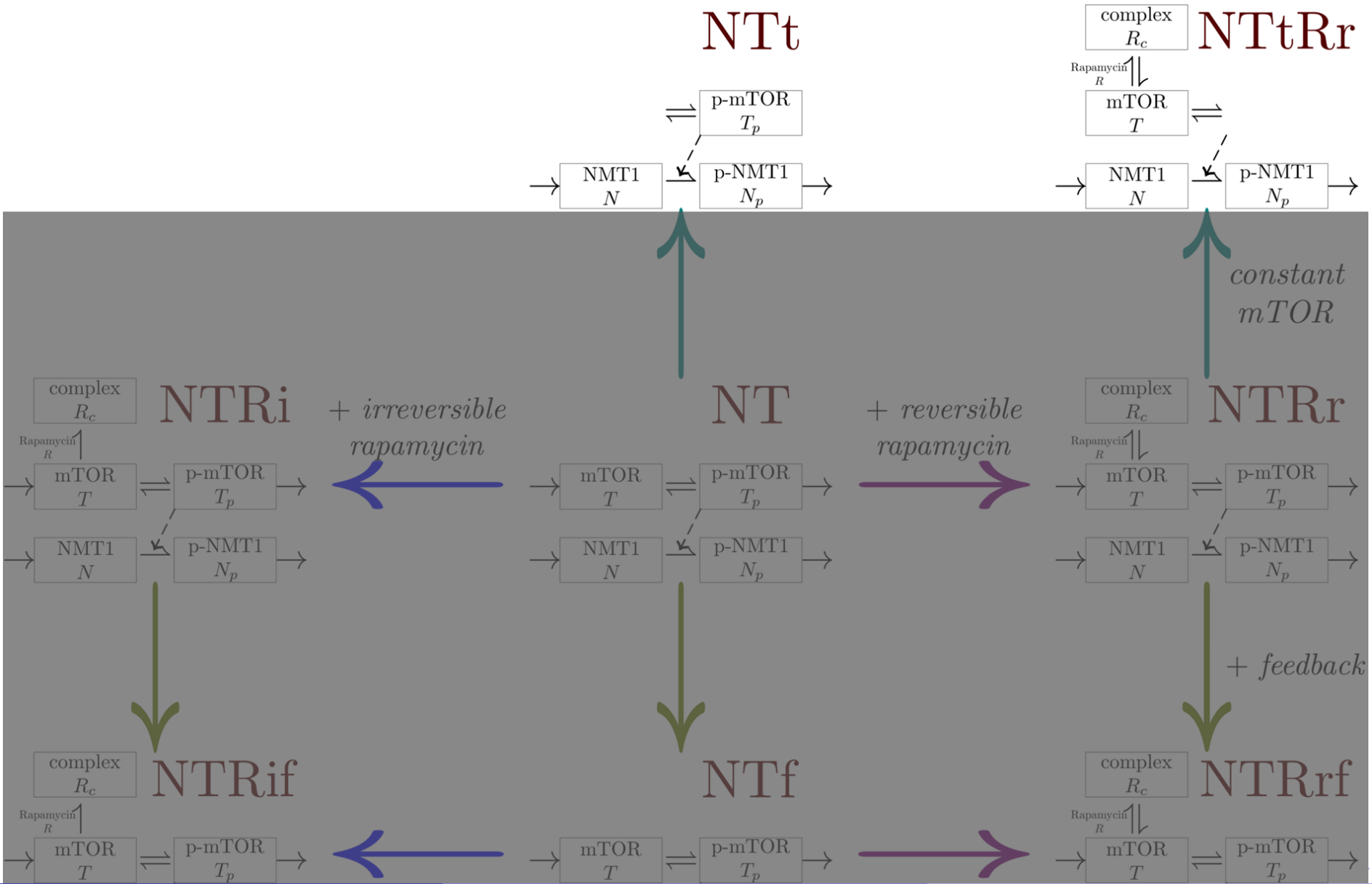
Overall: strong evidence of a constant endogenous level of total mTOR and an absence of negative feedback regulation of mTOR by NMT1

Proposed dynamical motif

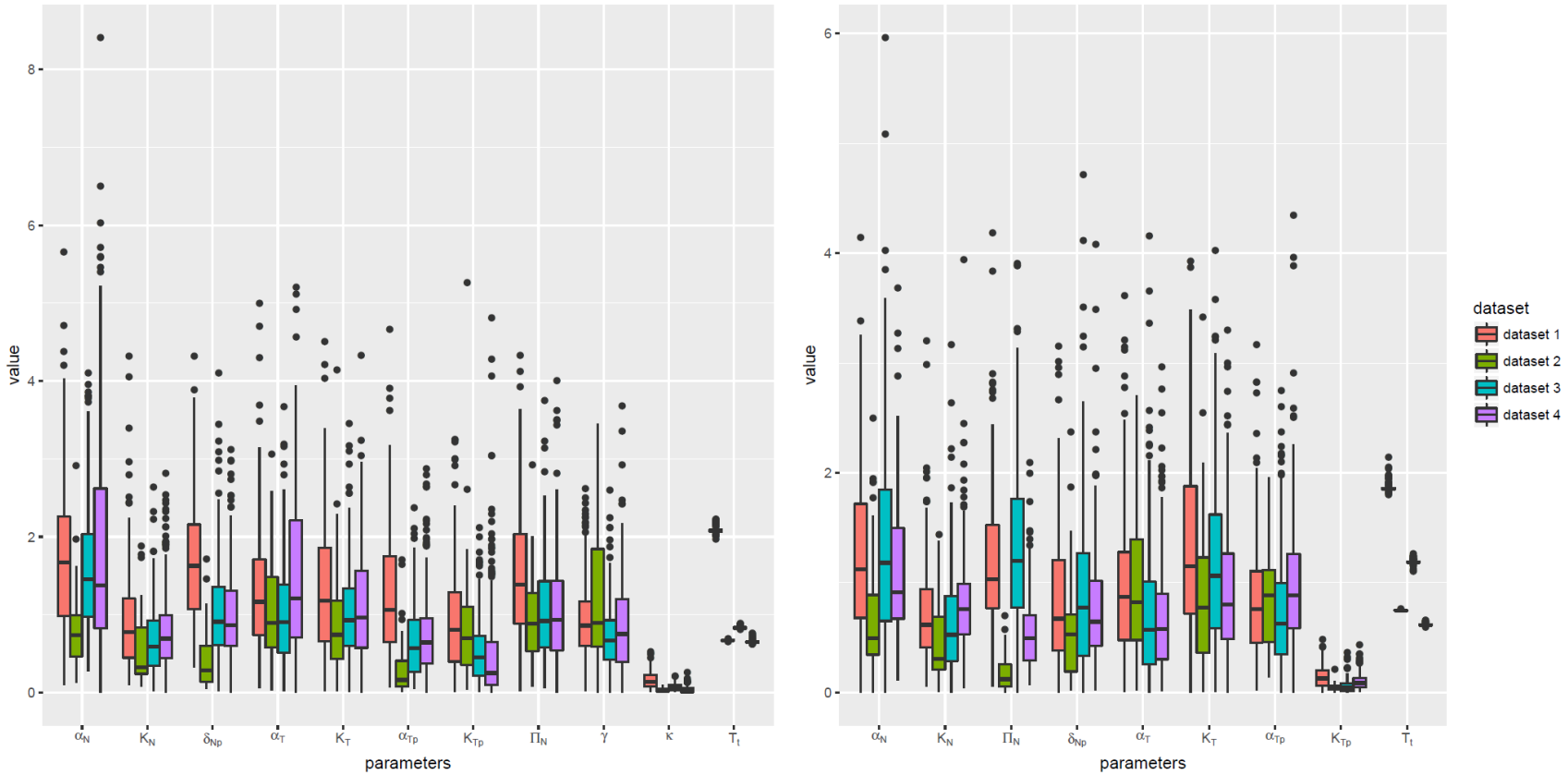
Perturbed system

Unperturbed system

Perturbed system



Parameters for the proposed dynamical motif



Perturbed system - NTtRr

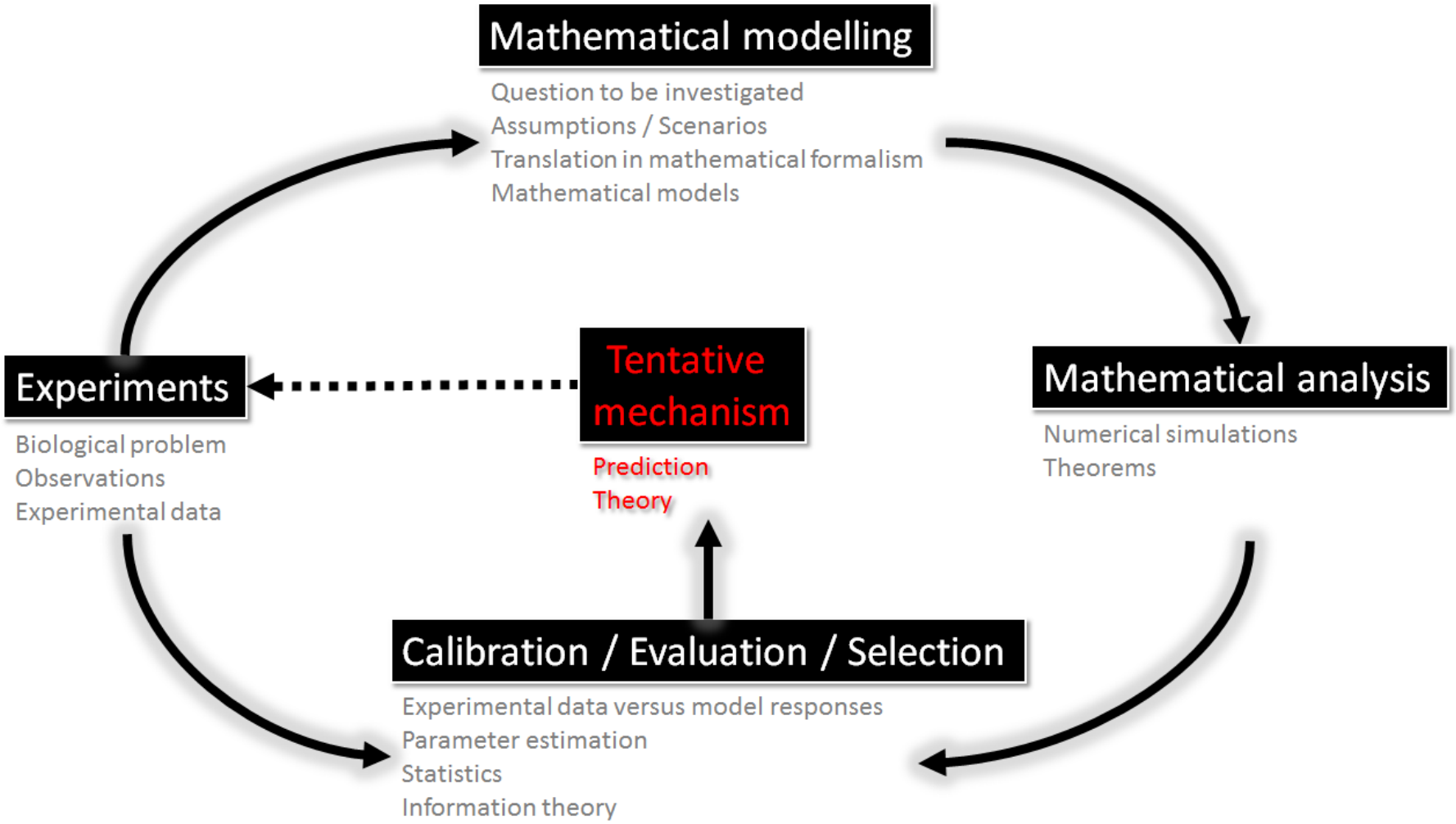
Unperturbed system - NTt

Depending on dataset used, different nominal sets of parameter values $\Rightarrow$ influence sensitivity analysis outcomes, prediction..

Outline

1. Combining data and mathematical models - how to get more
2. Example: Interactions between mTOR and NMT1 in breast cancer cells
3. Conclusion

Work flow – Mathematical modelling



Portet (2015) Insights E-Journal. 8(2).

My $0.02 on combining mathematical modelling and experimental data

To get more: Collection of models (with model selection):

- test different scenarios
- select model(s) that approximate the best data considered (Still not the truth!!)
- evaluate the relative importance of processes

To keep in mind: Conclusions drawn when combining mathematical modelling and data are impacted by

- mathematical translation of biological processes
- data considered

$\Rightarrow$ change in sensitivity analysis outcomes, transient (reactivity) and long term (prediction) dynamics...

Al-Darabsah, K.-L. Liao, and S. Portet, A simple in-host model for COVID-19 with treatments: model prediction and calibration.

Journal of Mathematical Biology, **86**:20 (2023)

Why use mathematical modelling?

- Test a large number of different scenarios
- Propose tentative hypotheses to be tested $\Rightarrow$ propose new experiments
- Identify the major components of processes
- Extrapolate the broad behavior of a system for which data cannot easily be obtained
- Theorize the processes (clarify hypotheses and characterize the chain of events)

“All models are wrong, but some are useful.” G. E. P. Box