Fusing Biophysics and Machine Learning for Computational Oncology

From tumor kinetics to therapy outcome

Cristian Axenie DrEngSc MSc BSc

AUDI KONFUZIUS-INSTITUT INGOLSTADT

Technische Hochschule Ingolstadt

Helios Klinikum München West
Akademisches Lehrkrankenhaus der Ludwig-Maximilians-Universität München

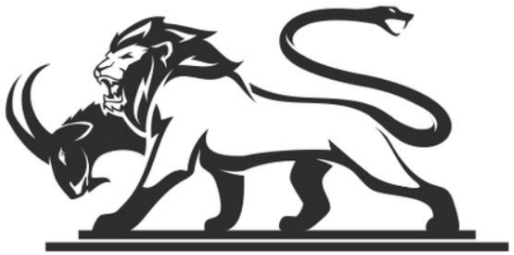
cristian.axenie@audi-konfuzius-institut-ingolstadt.de

Overview

A Framework for Mathematical and Computational Oncology

GLUECK

TUCANN



CHIMERA

PERFECTO

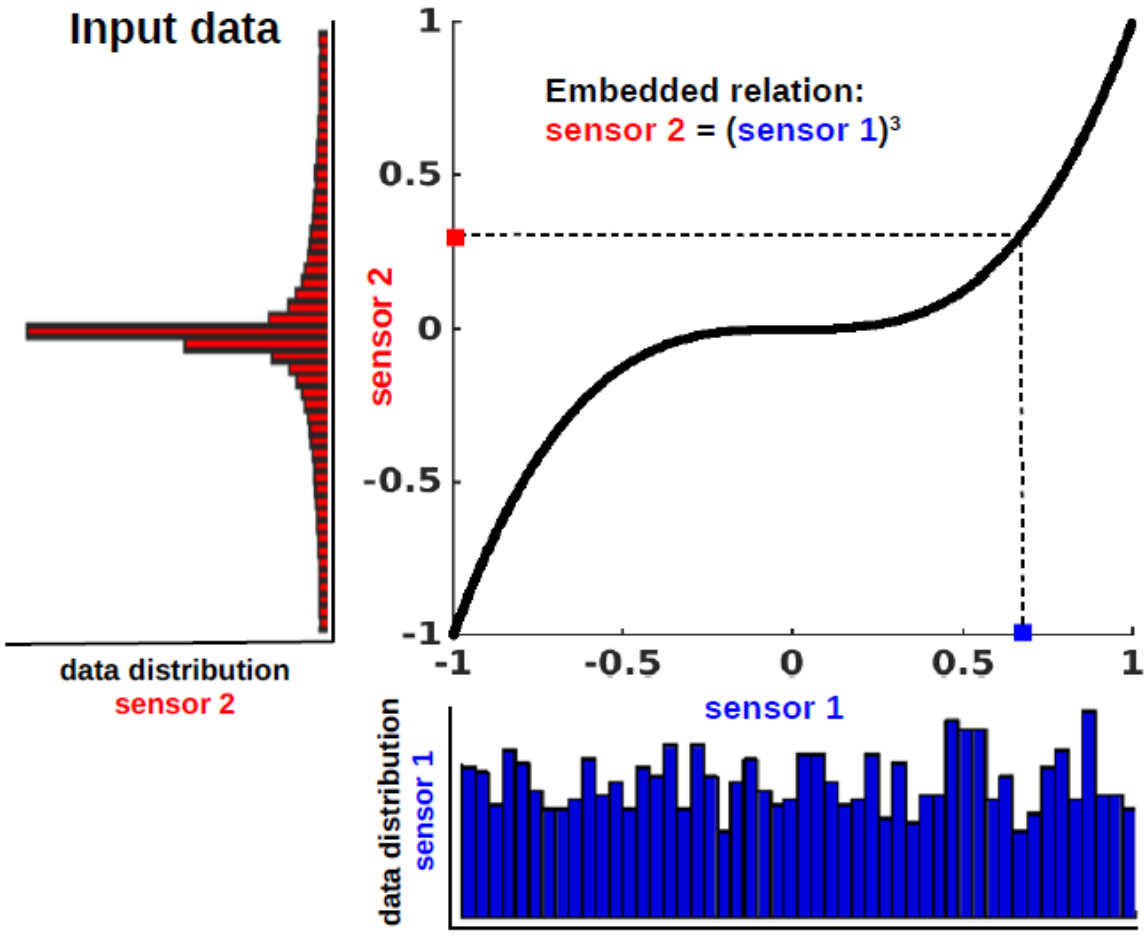
PRECISION

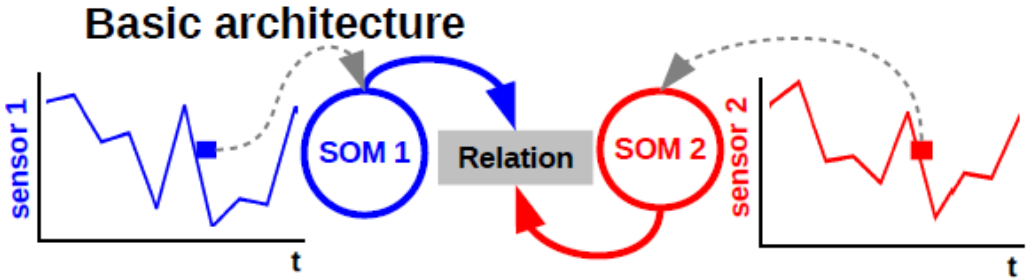
PROFILE EXTRACTION
FROM CLINICAL INSIGHTS
FOR SMART INDIVIDUALIZED ONCOLOGY

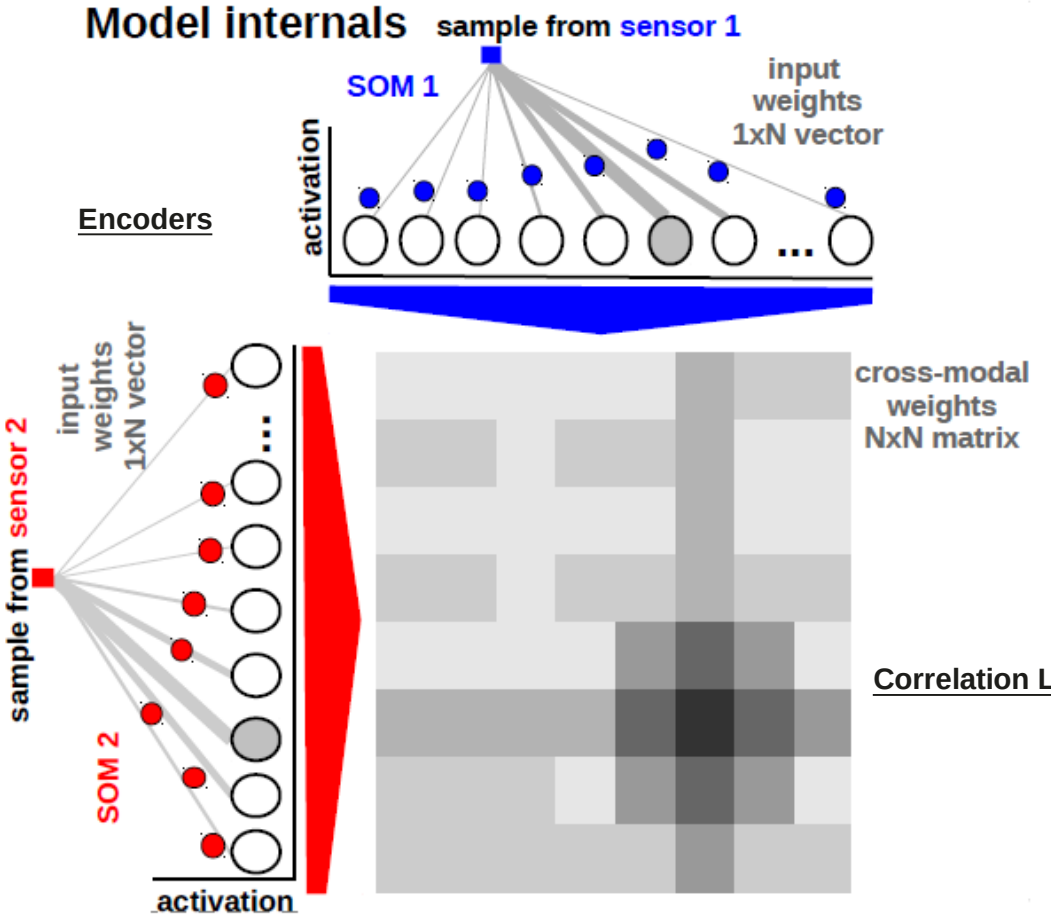
A Framework for Mathematical and Computational Oncology

Interacting Computational Maps

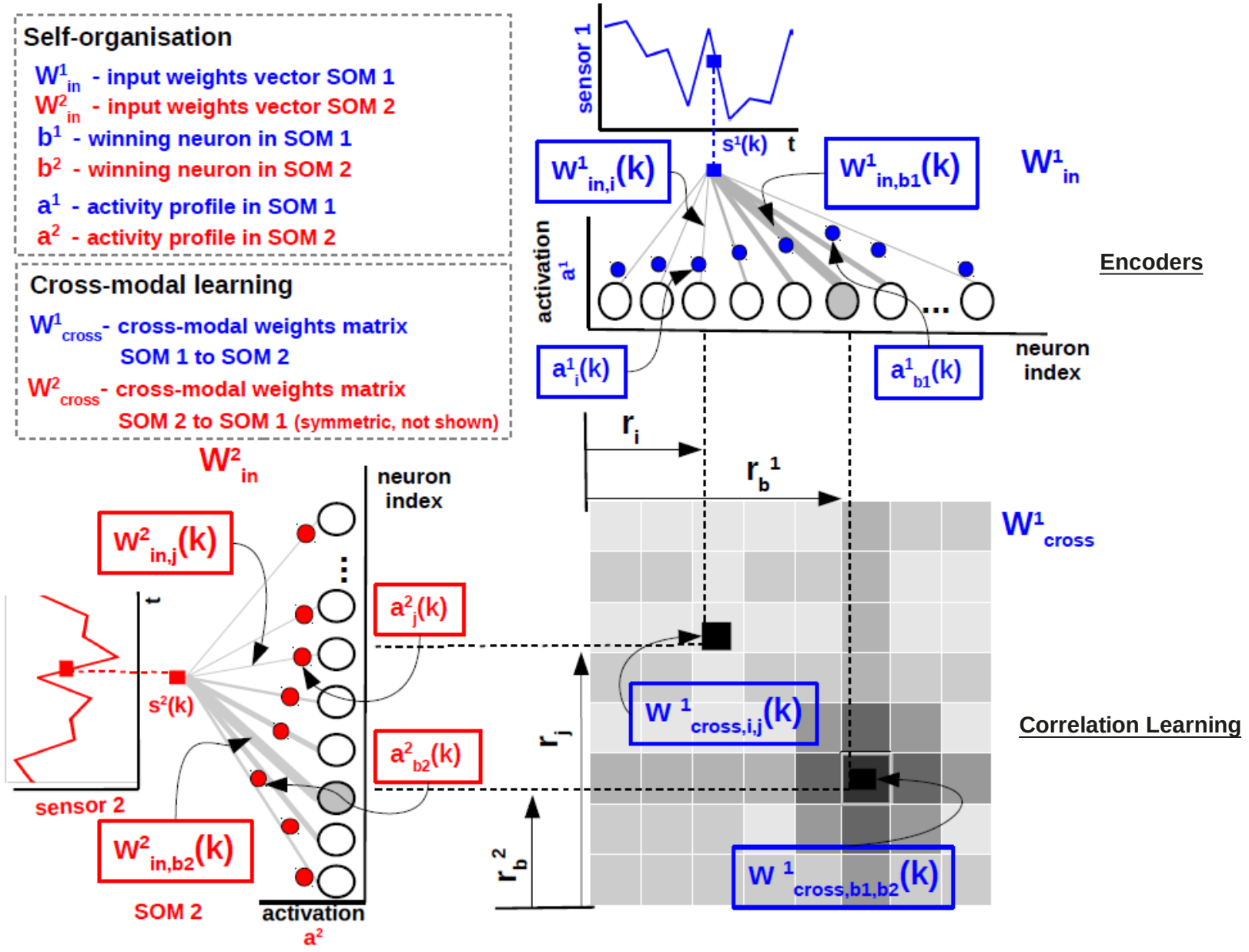
Core model


Input data
data distribution
sensor 2
Embedded relation:
sensor 2 = (sensor 1)³
sensor 2
sensor 1
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1
data distribution
sensor 1

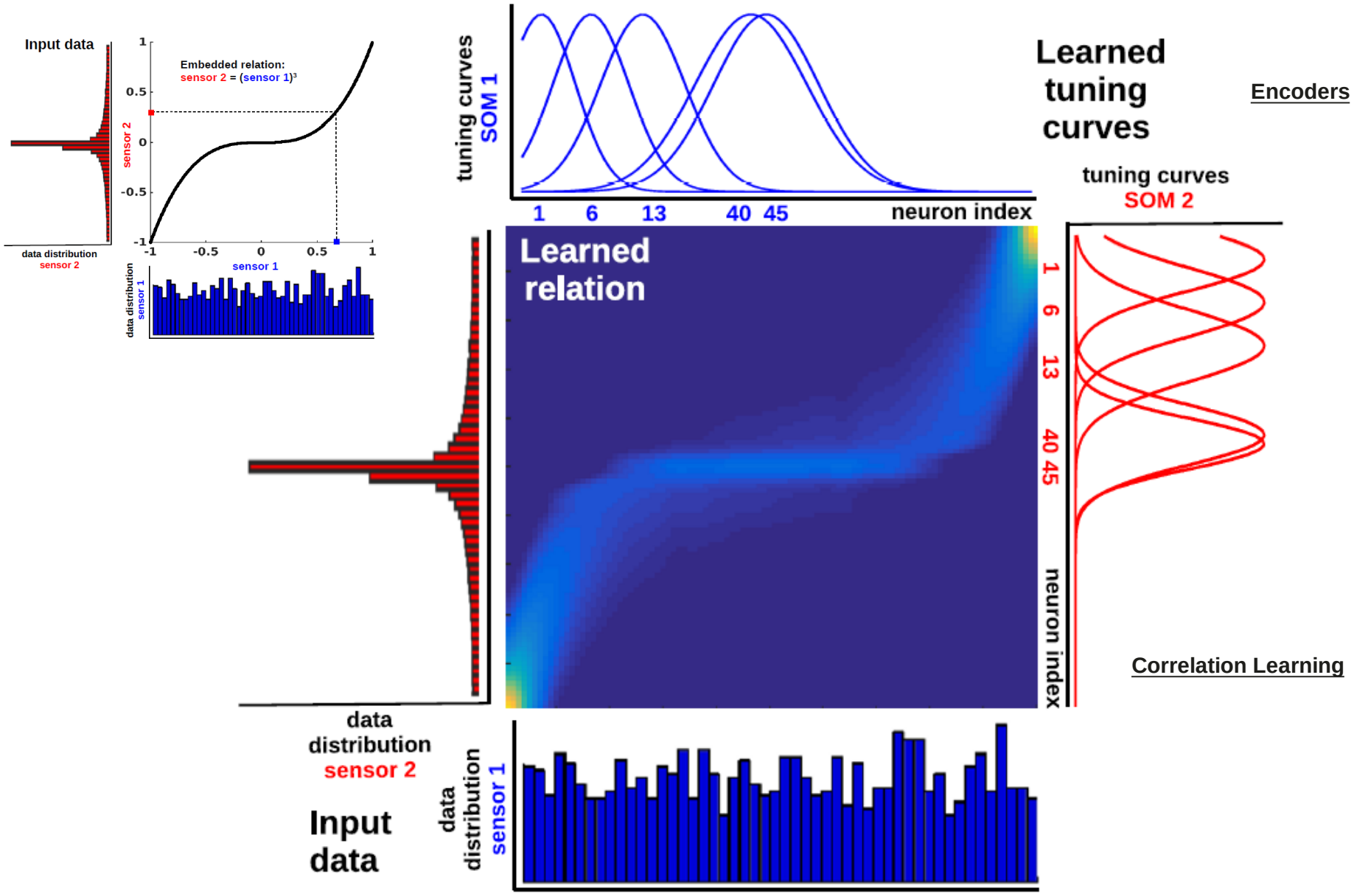

Basic architecture
sensor 1
t
SOM 1
Relation
SOM 2
sensor 2
t


Model internals
sample from sensor 1
SOM 1
input
weights
1xN vector
activation
Encoders
input
weights
1xN vector
sample from sensor 2
SOM 2
activation
cross-modal
weights
NxN matrix
Correlation Learning

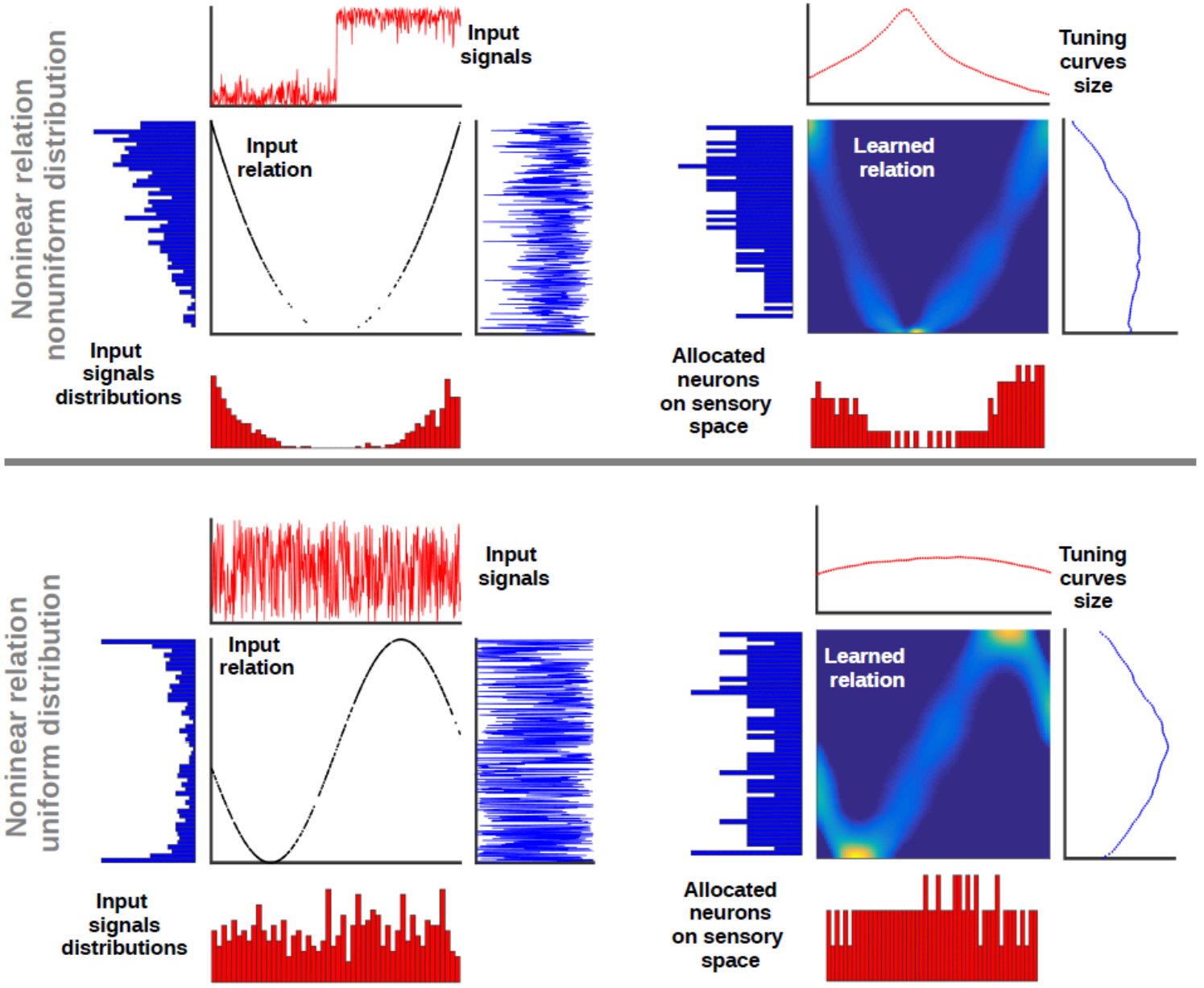
Core model internals



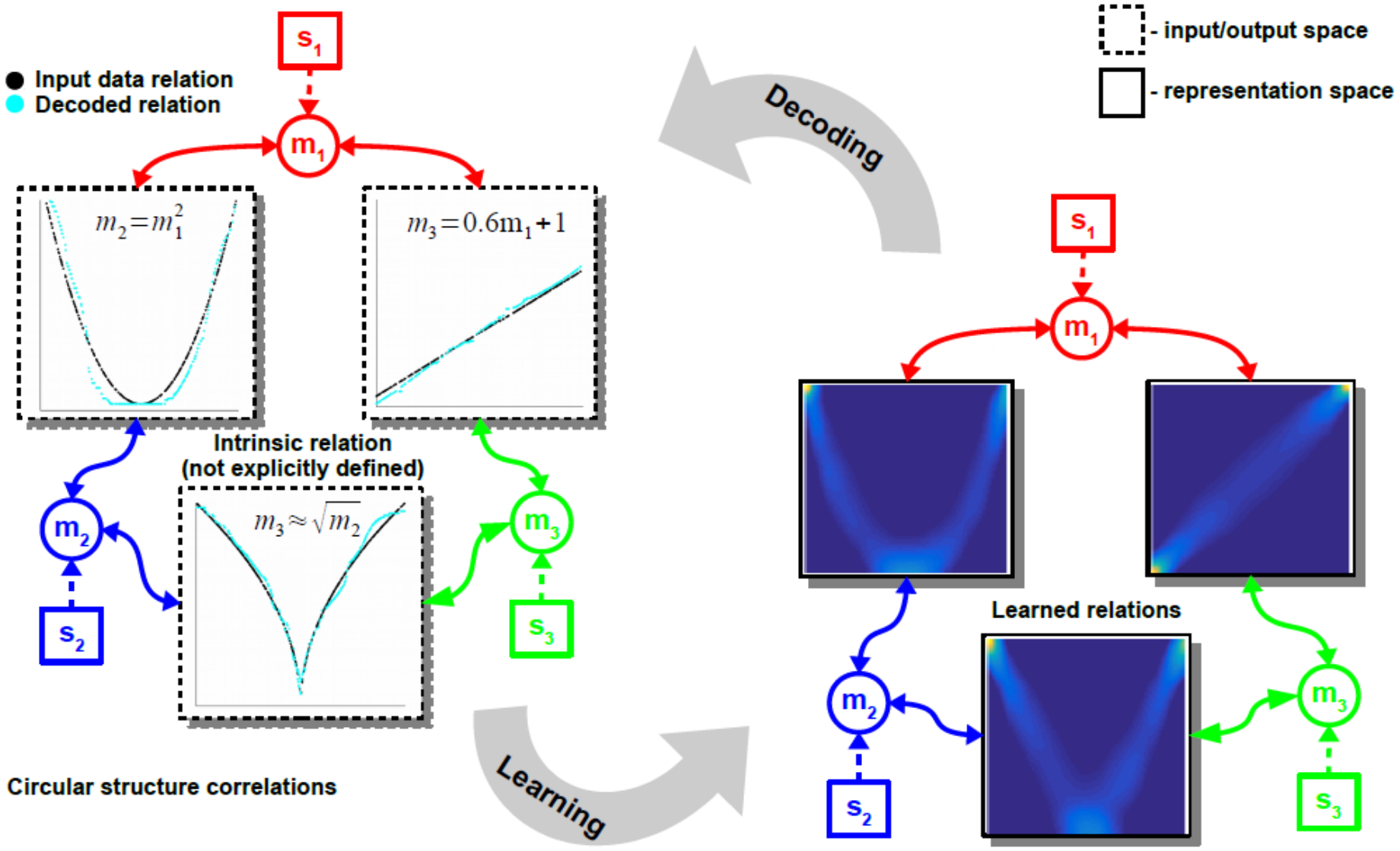
Learning capabilities I



Learning capabilities II



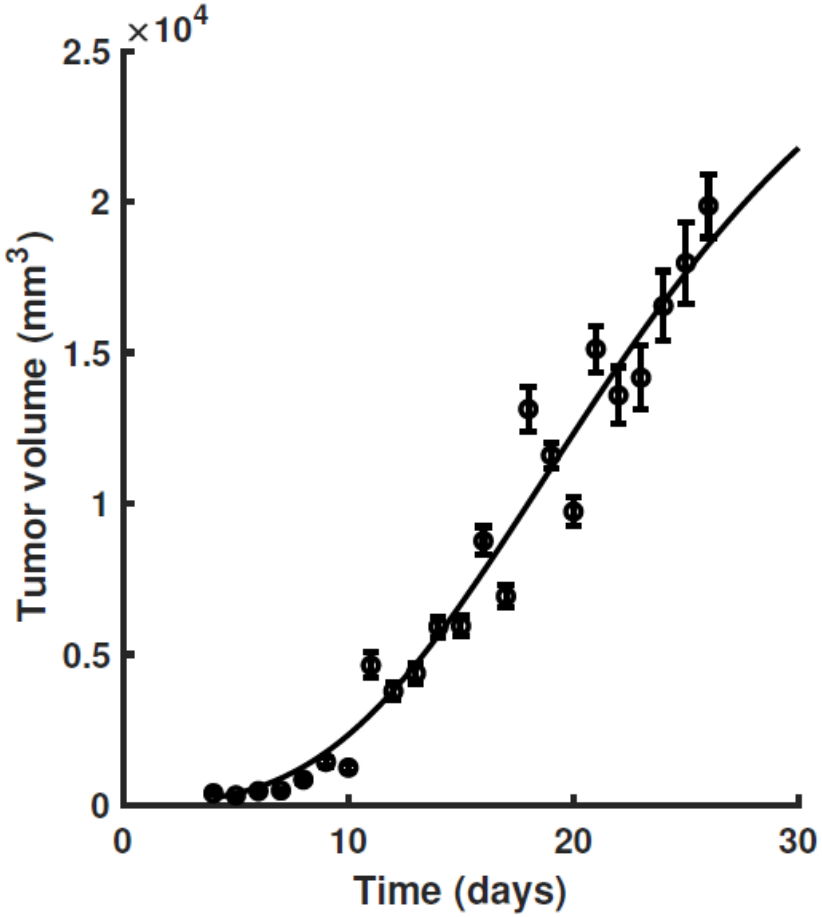
Extensibility

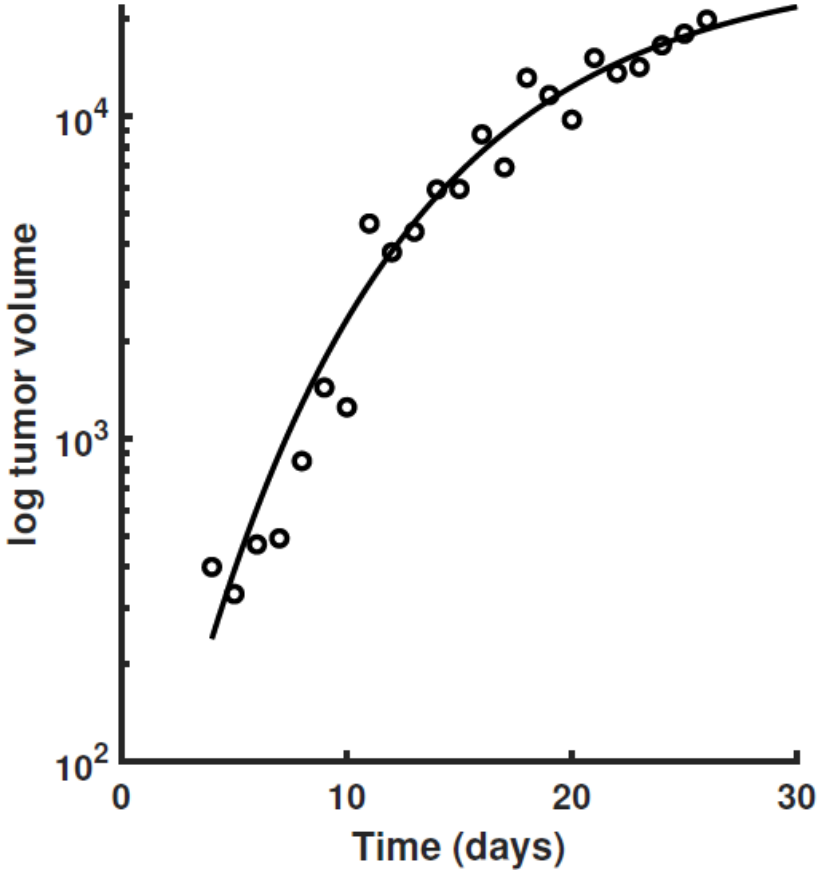


GLUECK

Growth pattern **L**earning for
Unsupervised **E**xtraction of **C**ancer **K**inetics

Tumor growth data



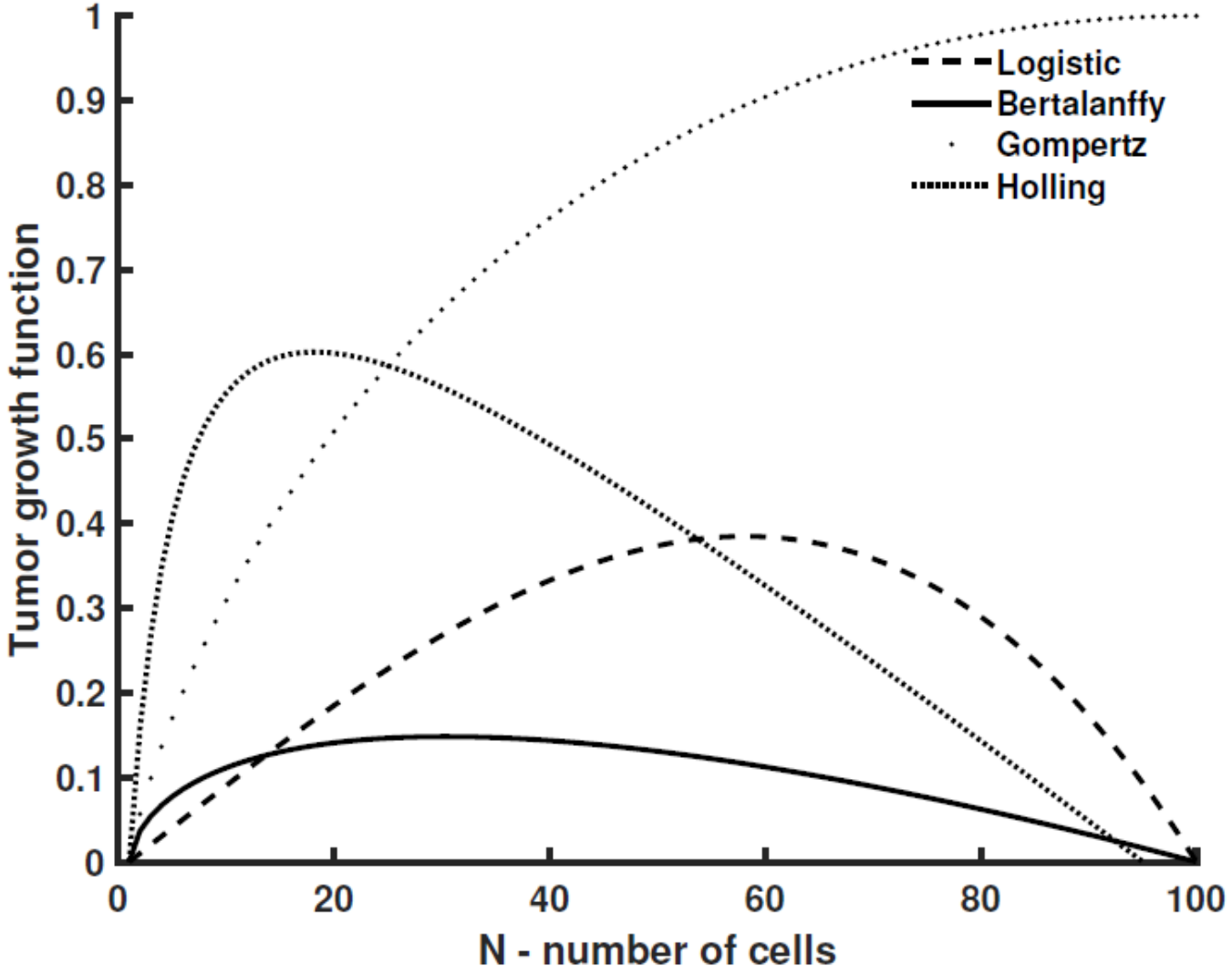


Peculiarities of data:

- Small
- Unevenly sampled
- High-variability
- Heterogeneous
- Model selection is hard
- Determines treatment

Growth kinetics of Fortner Plasmacytoma 1 tumors. Points represent mean volume of subcutaneous tumor implants in mice, error bars represent +/-1 standard error of the mean at each point. Data from Simpson-Herren et al. Cancer Chemotherapy Rep 54(3)

Tumor growth models

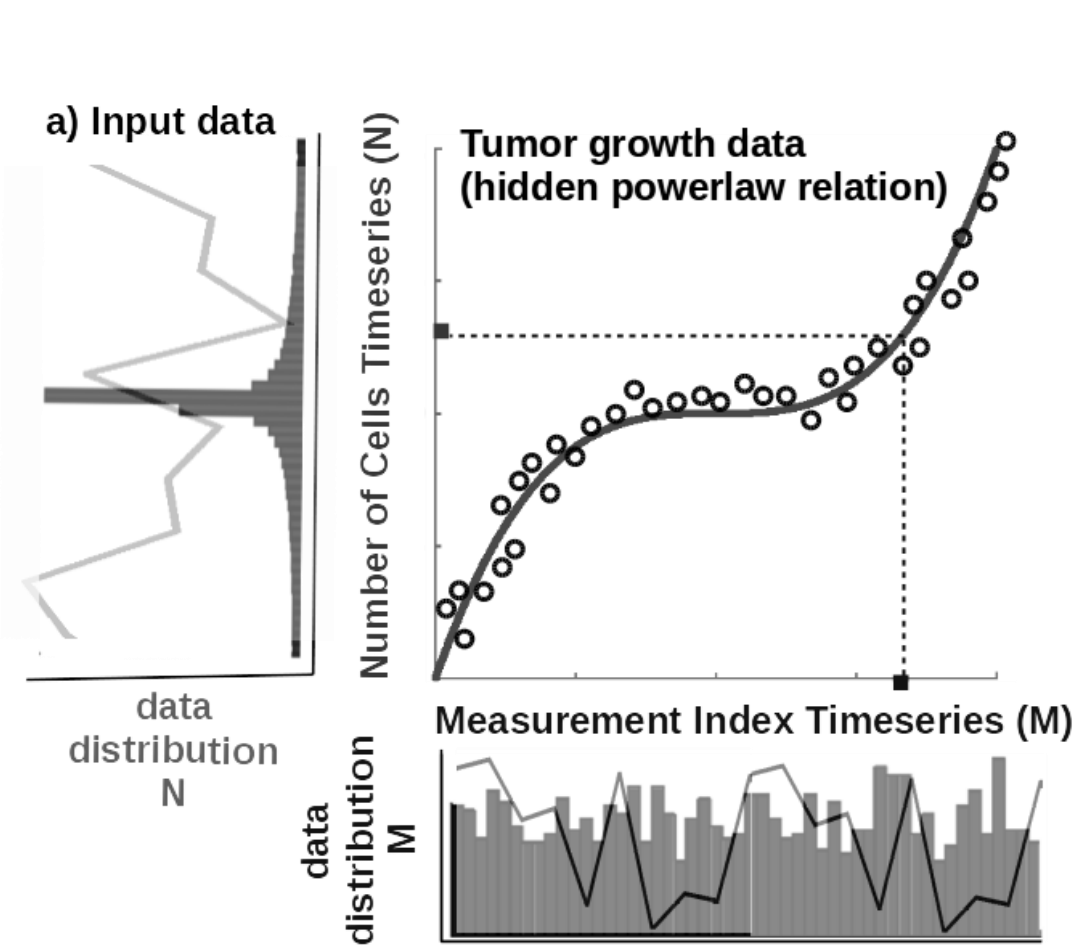


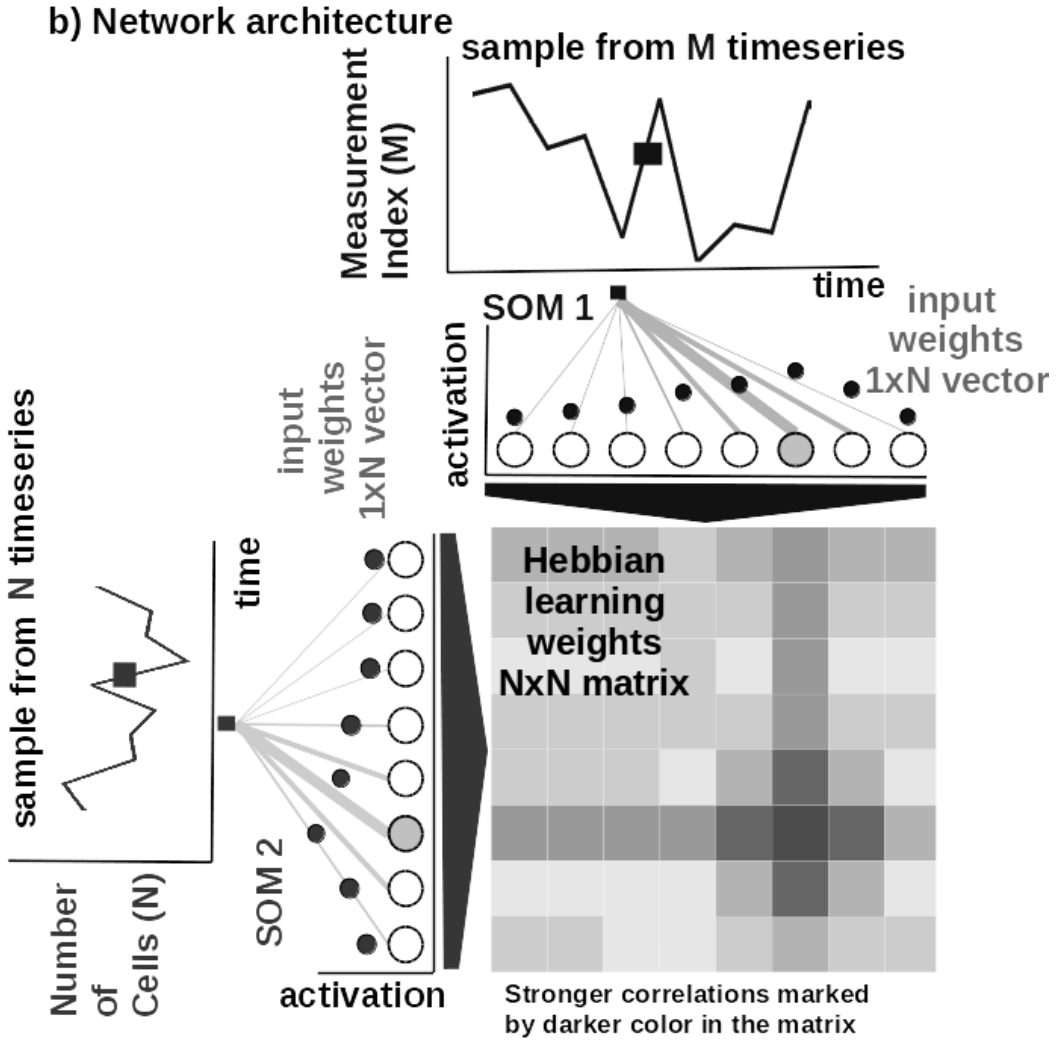
Model	Equation
Logistic	$\frac{dN}{dt} = \alpha N - \beta N^2$
Bertalanffy	$\frac{dN}{dt} = \alpha N^{\lambda} - \beta N$
Gompertz	$\frac{dN}{dt} = N(\beta - \alpha \ln N)$
Holling	$\frac{dN}{dt} = \frac{\alpha N}{k+N} - \beta N$

Parameters:

N - cell population size (or volume),
α - growth rate,
β - cell death rate,
λ - nutrient limited proliferation rate,
k - carrying capacity of cells.

Instantiating the model





Experiments and evaluation

Experimental dataset setup

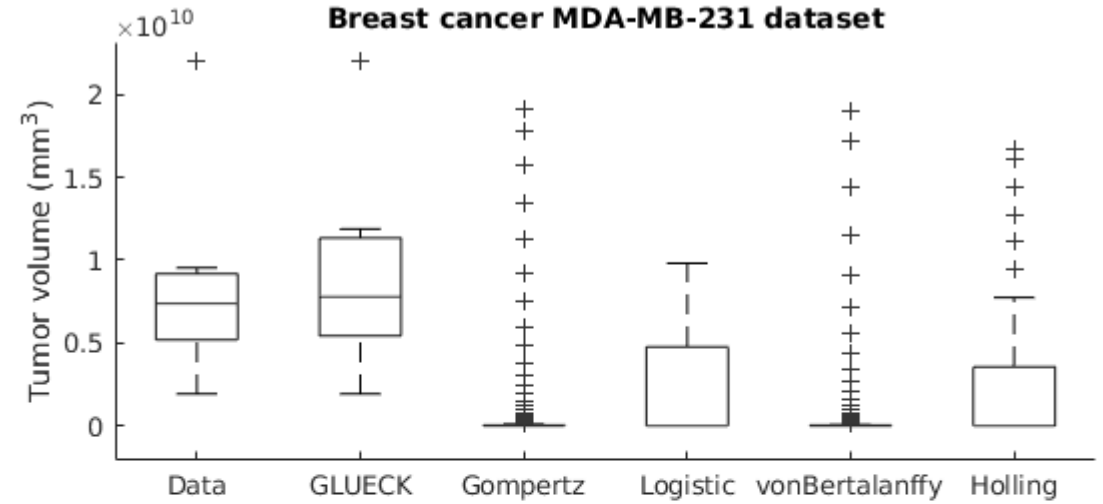
Dataset	Cancer Type	Data Type	Data Points	Data Freq.
1	Breast (MDA-MB-231)	Fluorescence imaging	7	2x/week
2	Breast (MDA-MB-435)	Digital Caliper	14	2x/week
3	Lung	Caliper	10	7x/week
4	Leukemia	Microscopy	23	7x/week

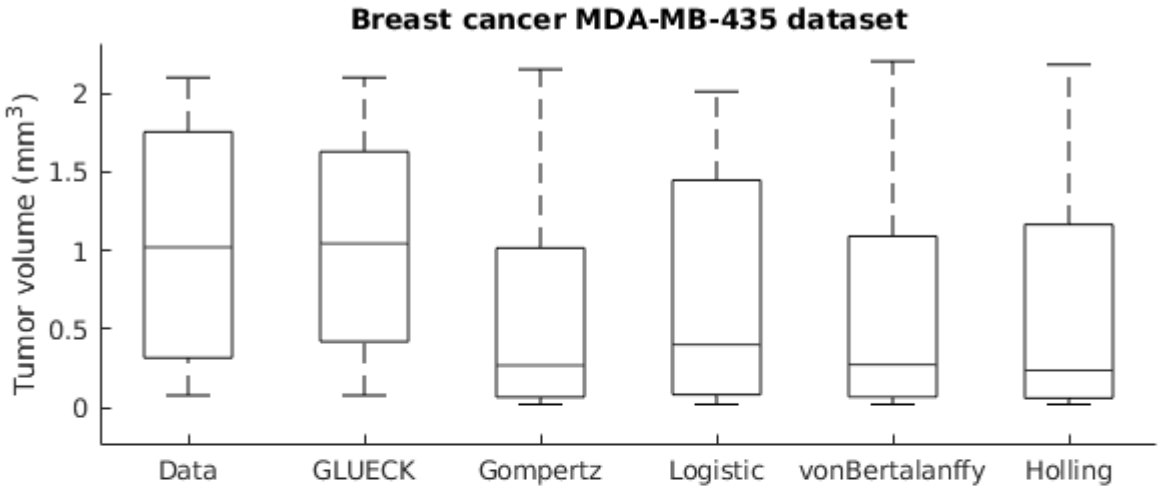
Metric	Equation
SSE	$\Sigma_{i=1}^{N}(\frac{y^i - y_m^i}{\sigma_i})$
RMSE	$\sqrt{\frac{SSE}{N-p}}$
sMAPE	$\frac{1}{N}\Sigma_{i=1}^{N}(2\frac{\lvert y^i - y_m^i \rvert}{(\lvert y^i \rvert + \lvert y_m^i \rvert)})$
AIC	$Nln(\frac{SSE}{N}) + 2p$
BIC	$Nln(\frac{SSE}{N}) + ln(N)p$

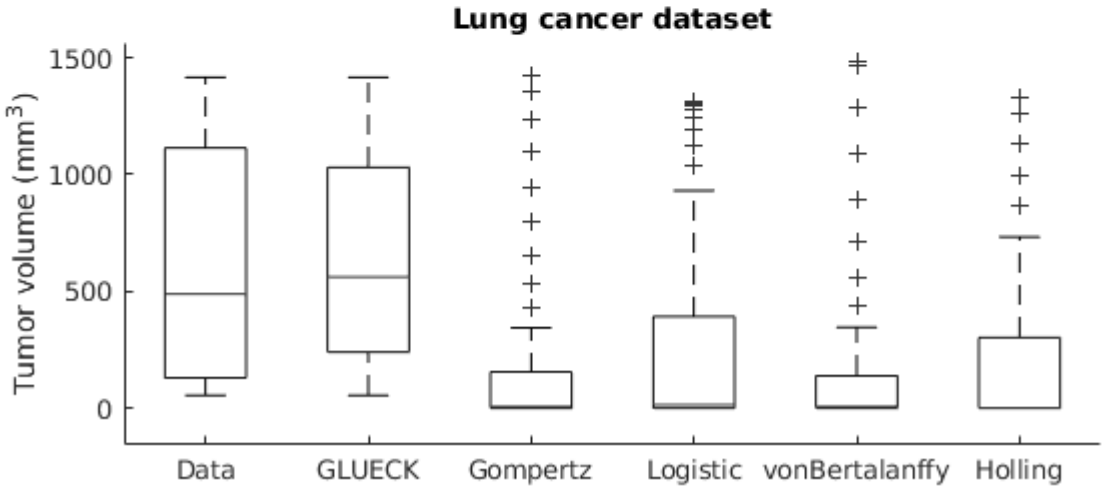
Evaluation metrics for tumor growth models.

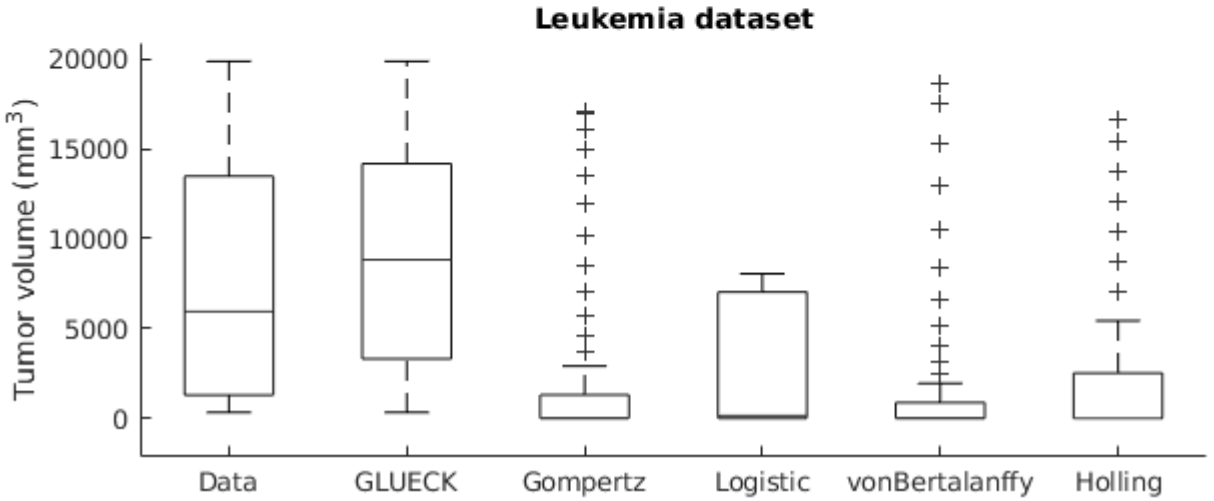
We consider:
N - number of measurements,
σ- standard deviation of data,
p - number of parameters of the model.

Experiments and evaluation









Experiments and evaluation

Dataset/Model	SSE	RMSE	sMAPE	AIC	BIC	Rank[a]
	Evaluation Metrics (smaller value is better)					
Breast[b] *cancer*[20]						
Logistic	7009.6	37.4423	1.7088	52.3639	52.2557	**2**
Bertalanffy	8004.9	44.7350	1.7088	55.2933	55.1310	**5**
Gompertz	7971.8	39.9294	1.7088	53.2643	53.1561	**4**
Holling	6639.1	40.7403	1.4855	53.9837	53.8215	**3**
GLUECK	119.3	4.1285	0.0768	19.8508	19.8508	**1**
Breast[c] *cancer*[26]						
Logistic	0.2936	0.1713	0.1437	-40.5269	-39.5571	**4**
Bertalanffy	0.2315	0.1604	0.1437	-41.3780	-39.9233	**2**
Gompertz	0.3175	0.1782	0.1437	-39.5853	-38.6155	**5**
Holling	0.2699	0.1732	0.1512	-39.5351	-38.0804	**3**
GLUECK	0.0977	0.0902	0.0763	-57.7261	-57.7261	**1**

Dataset/Model	SSE	RMSE	sMAPE	AIC	BIC	Rank[a]
	Evaluation Metrics (smaller value is better)					
Lung cancer[6]						
Logistic	44.5261	2.2243	1.5684	19.3800	20.1758	**2**
Bertalanffy	54.1147	2.6008	1.5684	23.5253	24.7190	**5**
Gompertz	53.2475	2.4324	1.5684	21.3476	22.1434	**4**
Holling	50.6671	2.5166	1.5361	22.8012	23.9949	**3**
GLUECK	3.6903	0.5792	0.2121	-12.0140	-12.0140	**1**
Leukemia[23]						
Logistic	223.7271	3.2640	1.6368	56.3235	58.5944	**2**
Bertalanffy	273.6770	3.6992	1.6368	62.9585	66.3649	**5**
Gompertz	259.9277	3.5182	1.6368	59.7729	62.0439	**4**
Holling	248.5784	3.5255	1.6001	60.7461	64.1526	**3**
GLUECK	35.2541	1.2381	0.3232	9.8230	9.8230	**1**

[a] Calculated as best in 3/5 metrics.
[b] MDA-MB-231 cell line
[c] MDA-MB-435 cell line

TUCANN

TUmor Phenotypical Transitions **C**haracterization using **A**rtificial **N**eural **N**etworks

Phenotypical transitions of tumors in DCIS

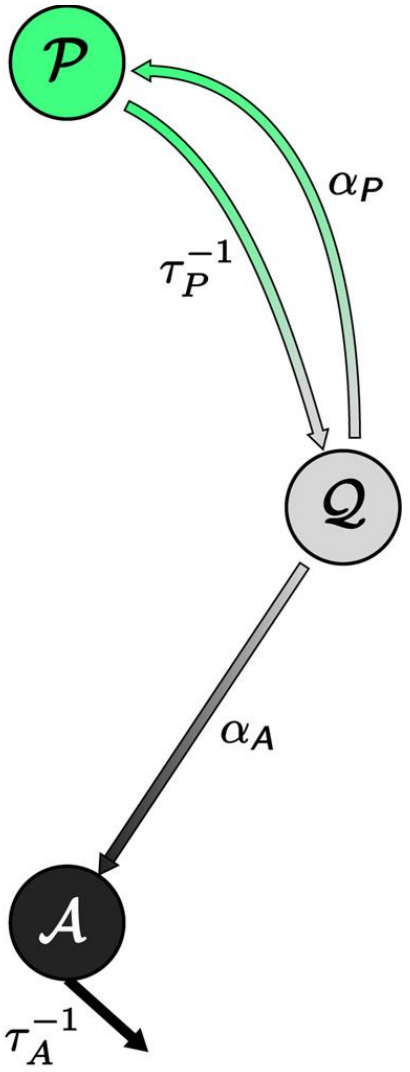
In typical cancer **phenotypic state space**, **quiescent** cancer cells (**Q**) can become **proliferative** (**P**) or **apoptotic** (**A**).

Can we learn **phenotypical transitions** from timeseries of raw immunohistochemistry and morphometric data?

$$\alpha_P = \frac{\frac{1}{\tau_P}(PI + PI^2) - \frac{1}{\tau_A} AIPI}{1 - AI - PI}$$

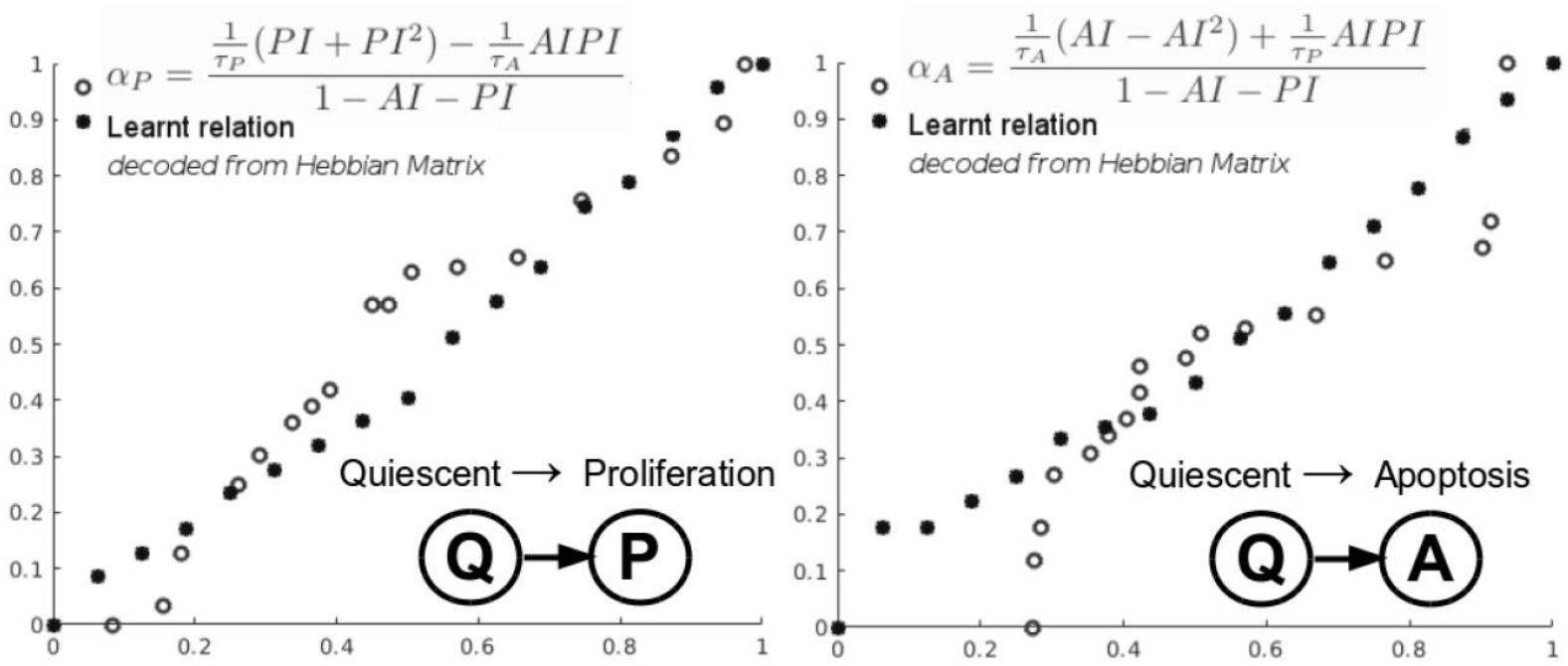
$$\alpha_A = \frac{\frac{1}{\tau_A}(AI - AI^2) + \frac{1}{\tau_P} AIPI}{1 - AI - PI}$$

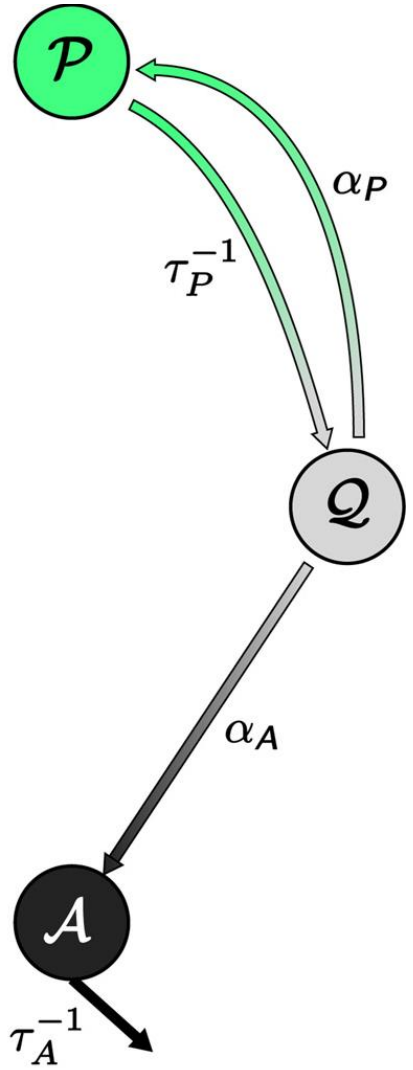
where, τ_P is the cells cycle time, τ_A cells apoptosis time , PI proliferation index and AI apoptosis index.

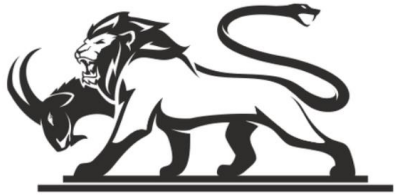


Phenotypical transitions of tumors in DCIS

Can we learn phenotypical transitions from timeseries of raw immunohistochemistry and morphometric data?







CHIMERA

Combining Mechanistic Models and Machine Learning for Chemotherapy-Surgery Sequencing

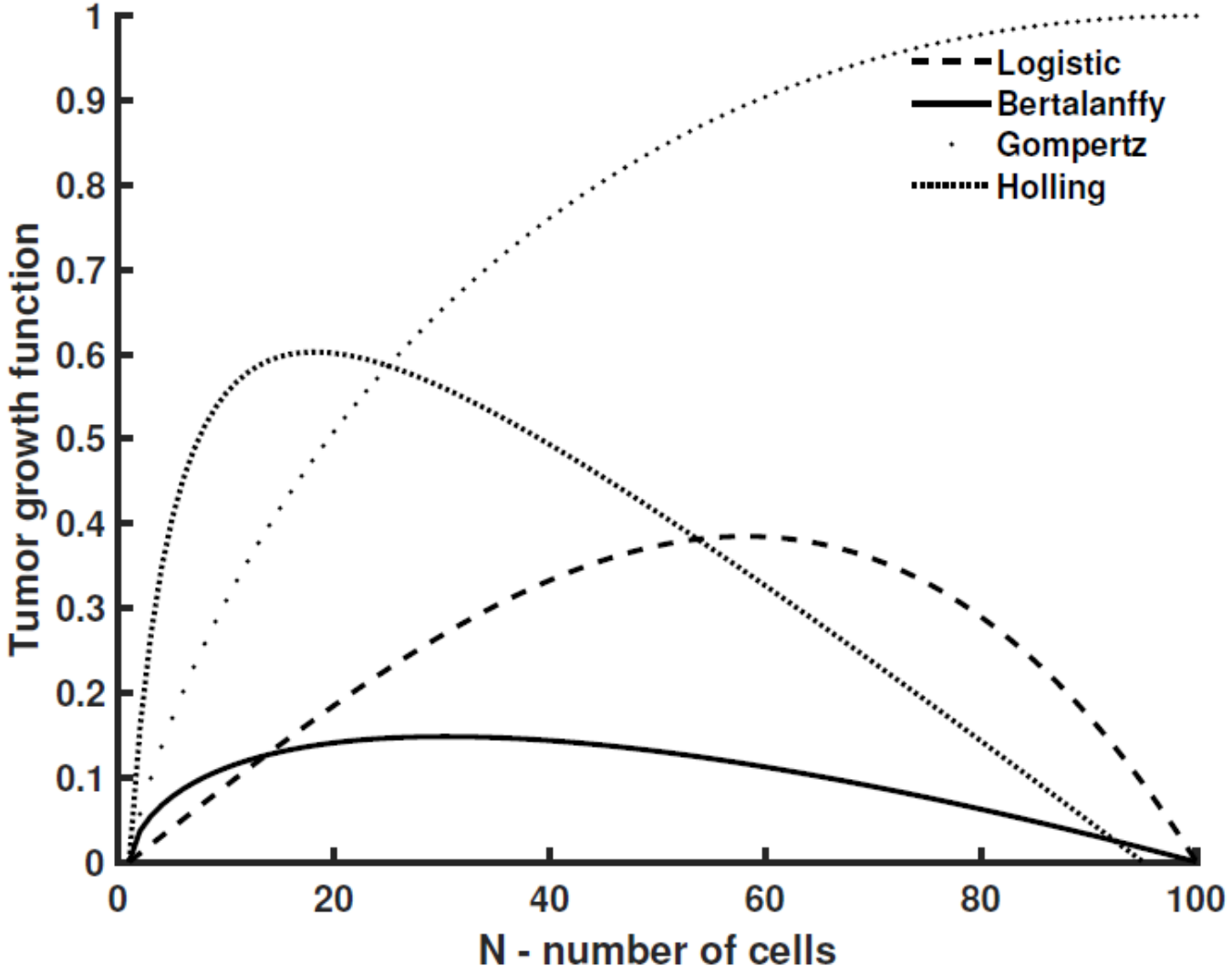
Formalizing therapy sequencing

If we consider $\boldsymbol{f(V)}$ the **tumor growth** model and $\boldsymbol{P(t,V)}$ **the pharmacokinetics** of the chemotherapeutic drug, we can formalize the two sequences as following:

- **Sequence 1**: **Adjuvant** setting, where size before surgery is $\frac{dv_1}{dt} = f(v_1), v_1(0) = V_0, t \in [0, t_0]$ and size after surgery is
$$\frac{dV_1}{dt} = f(V_1) - P(t, V_1), V_1(t_0) = e^{-k_s} v_1(t_0), t \in [t_0, t_f].$$
In this case, the final volume of the tumor is $V_{adj} = V_1(t_f)$.

- **Sequence 2**: **Neoadjuvant** setting, where the size before chemotherapy onset is $\frac{dv_2}{dt} = f(v_2), v_2(0) = V_0, t \in [0, t_0]$ and the size after chemotherapy onset is
$$\frac{dV_2}{dt} = f(V_2) - P(t, V_2), V_2(t_0) = v_2(t_0), t \in [t_0, t_f]$$ respectively.
Hence, for the neoadjuvant sequence, the final volume of the tumor is
$V_{neoadj} = e^{-k_s} V_2(t_f)$.

Tumor growth models



Model	Equation
Logistic	$\frac{dN}{dt} = \alpha N - \beta N^2$
Bertalanffy	$\frac{dN}{dt} = \alpha N^{\lambda} - \beta N$
Gompertz	$\frac{dN}{dt} = N(\beta - \alpha \ln N)$
Holling	$\frac{dN}{dt} = \frac{\alpha N}{k+N} - \beta N$

Parameters:

N - cell population size (or volume),
α - growth rate,
β - cell death rate,
λ - nutrient limited proliferation rate,
k - carrying capacity of cells.

Pharmacokinetics models

In our study, we use the data from the computational model of **Paclitaxel pharmacokinetics** of Kuh et al. 2000 [8], due to its wide use in **breast cancer chemotherapy schemes**.

The model describes the factors that determine the kinetics of **Paclitaxel uptake**, **binding**, and **efflux** from cells

$$\frac{dc(t)}{dt} = \left[\frac{-A + \sqrt{A^2 + 4K_{d,m}c_m(t)}}{2} - \frac{-B + \sqrt{B^2 + 4(1+NSB)K_{d,c}c\,(t)}}{2(1+NSB)}\right]\frac{CL_f}{V_{onecell}} - k_{cellnumber}c(t)$$

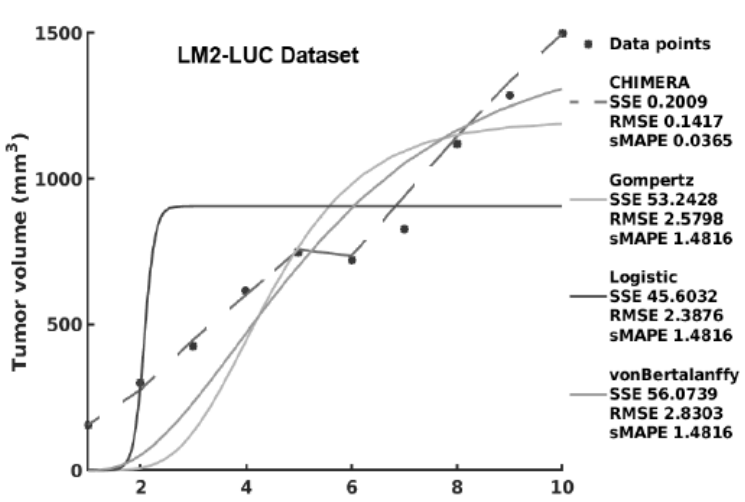
where:

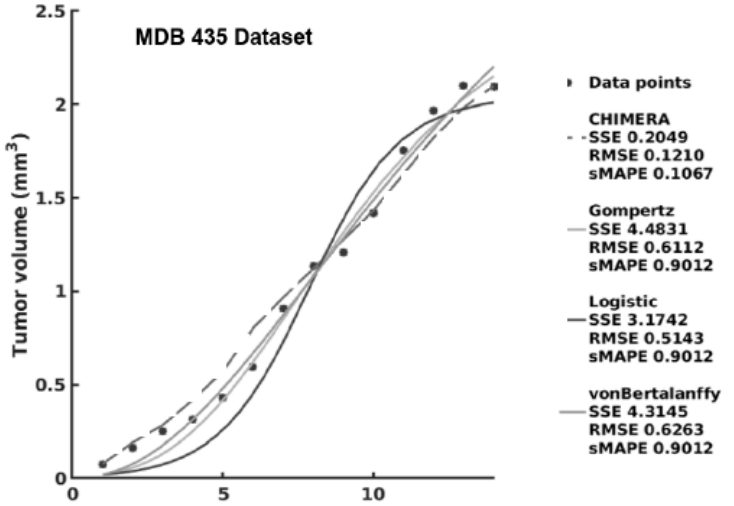
- $V_{onecell}$ is the average cell volume
- ICN is the initial cell number
- NSB is the proportionality constant for non-saturable binding sites in cells
- $k_{cellnumber}$ is the rate constant for changes in cell number
- A is a function of the constant for drug binding to proteins in medium $K_{d,m}$
- B is a function of the constant for drug binding to proteins in cells
- CL_f is the clearance of free drug by passive diffusion, on a per cell basis
- c_m concentration of drug in the medium, calculated as:

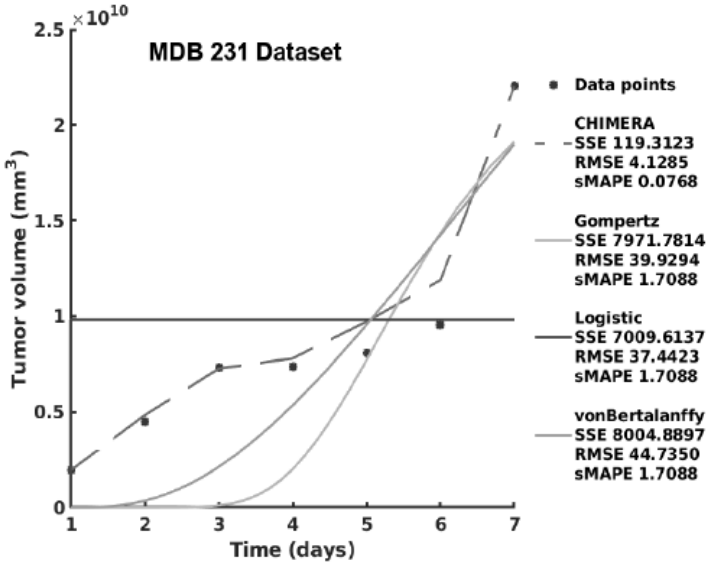
$$\frac{dc_m(t)}{dt} = \left[\frac{-A + \sqrt{A^2 + 4K_{d,m}c_m(t)}}{2} - \frac{-B + \sqrt{B^2 + 4(1+NSB)K_{d,c}c\,(t)}}{2(1+NSB)}\right]\frac{CL_f ICN e^{k_{cellnumber}t}}{V_m}$$

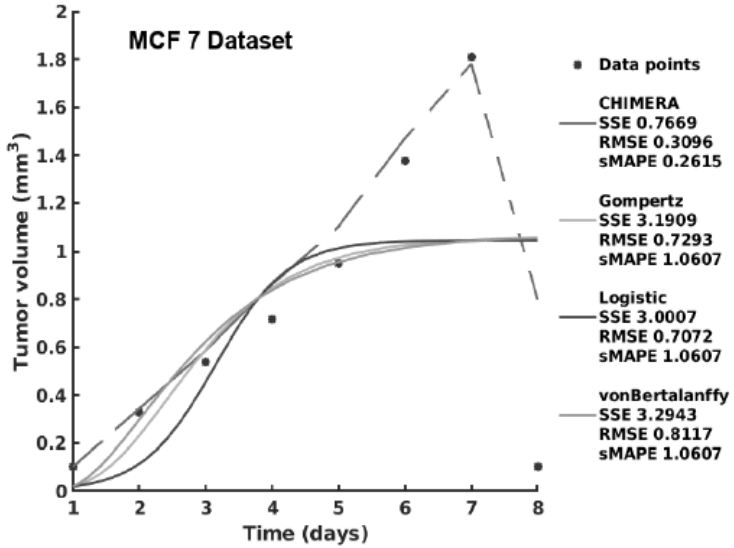
Experiments and evaluation

Learning tumor growth



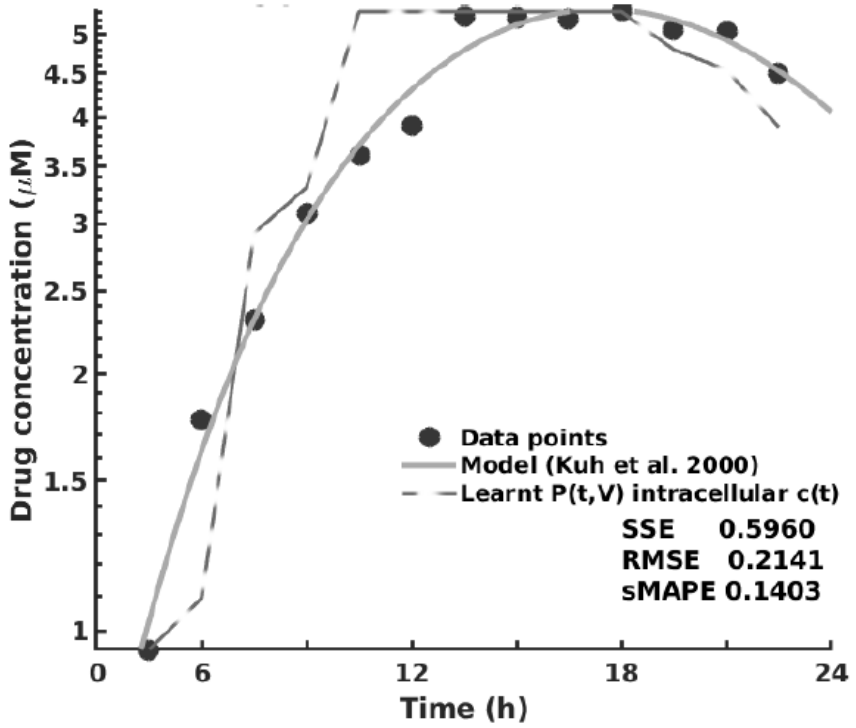


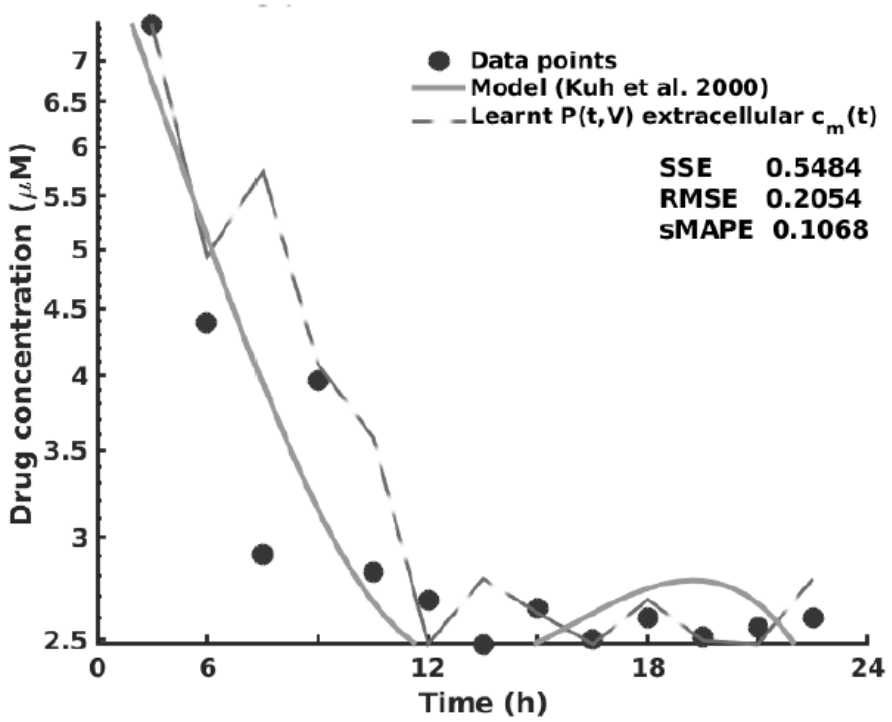




Experiments and evaluation

Learning pharmacokinetics





Experiments and evaluation

Chemotherapy-Surgery Sequencing

Let's consider the **MCF-7 cell line dataset** from Tan et al. 2015 [16] described in our Experimental setup. We use the derivations for V_{neoadj} and V_{adj} and fill in with the **decoded values** from the **learnt tumor growth** $f(V)$ and **learnt pharmacokinetics** $P(t, V)$.

Model (Biological Parameters)	Log-kill hypothesis	Norton-Simon hypothesis
Gompertz (β, K, υ)	$V_{neoadj} < V_{adj}$	$V_{neoadj} < V_{adj}$ *
CHIMERA (none)	$V_{neoadj} < V_{adj}$	$V_{neoadj} > V_{adj}$

* Holds only if $c(t) = \int_{t_0}^{t_f} c(s)ds < t_f - t_0$.

CHIMERA uses **learnt tumor growth and pharmacokinetics** to **infer** the most appropriate **sequence of therapy**, **consistent with its mechanistic counterparts**, but **without** extensive biological **parametrization**.

PERFECTO

Prediction of **E**xtended **R**esponse and Growth
Functions for **E**stimating **C**hemo**t**herapy **O**utcomes

Chemotherapy regimen planning

Context

Chemotherapy regimens are chosen primarily based on:

- empirical **data** from **clinical trials**
- patient's **form** and **subtype** of **cancer**
- **progression** to metastases
- **high-risk indications**
- **prognosis**

Problem

Challenges in successfully **predicting the effectiveness** (i.e. size of the tumor after **neoadjuvant chemotherapy**) of any particular chemotherapy plan for any given patient **before the initiation of the regimen**.

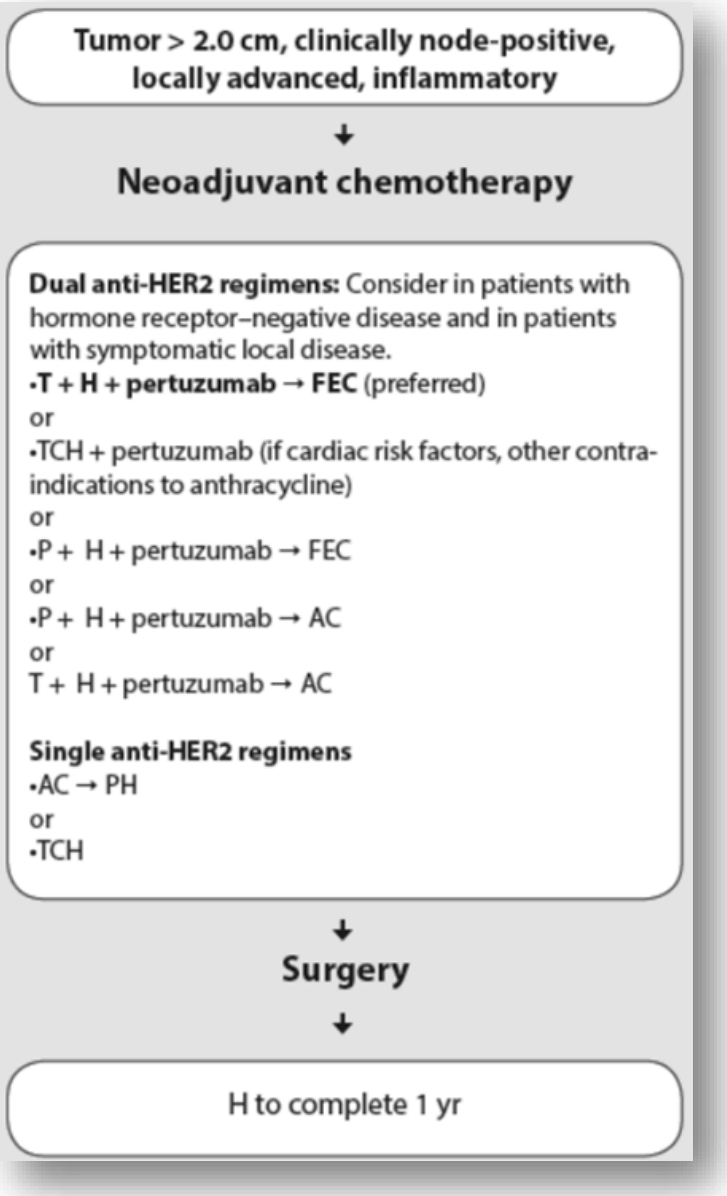
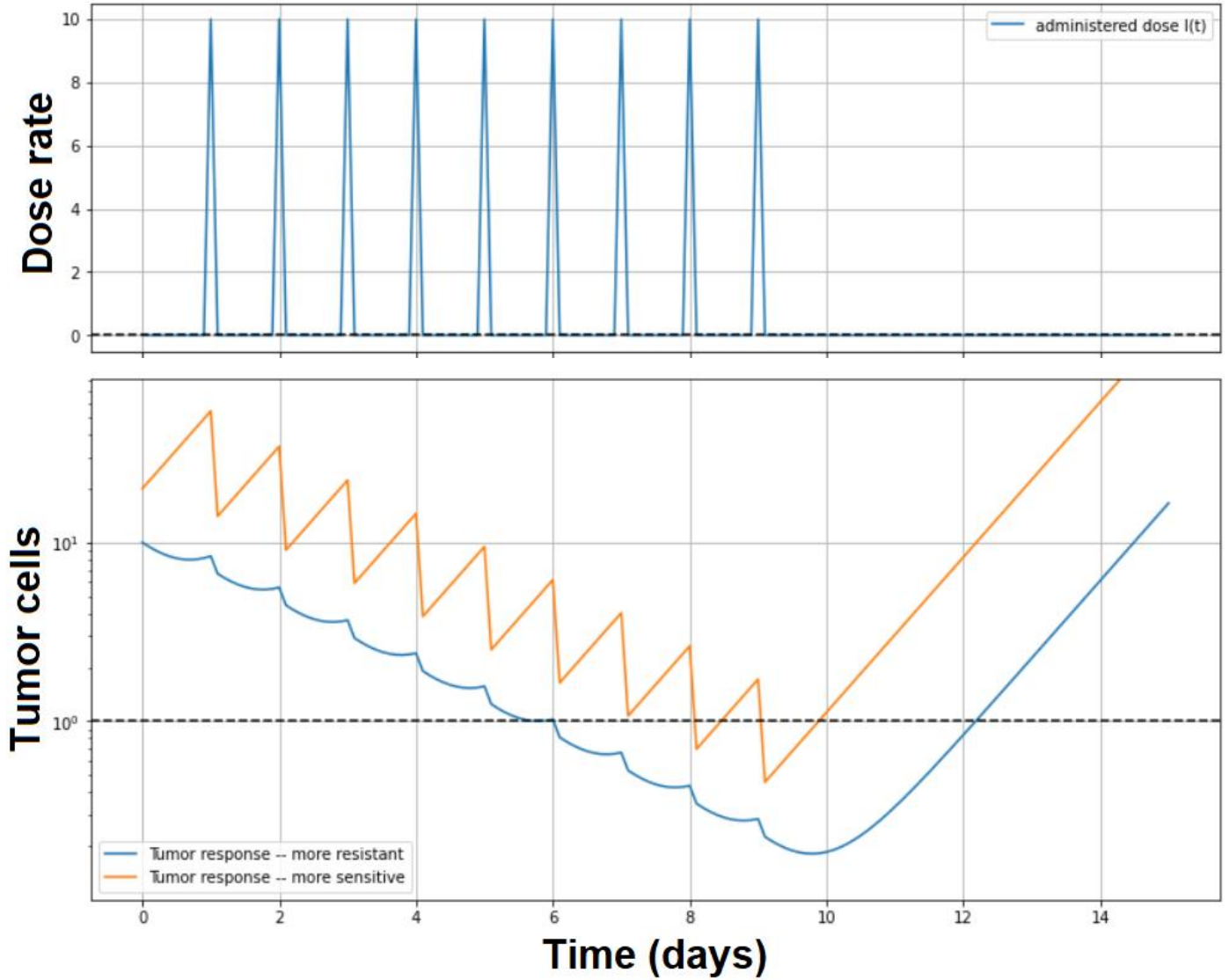


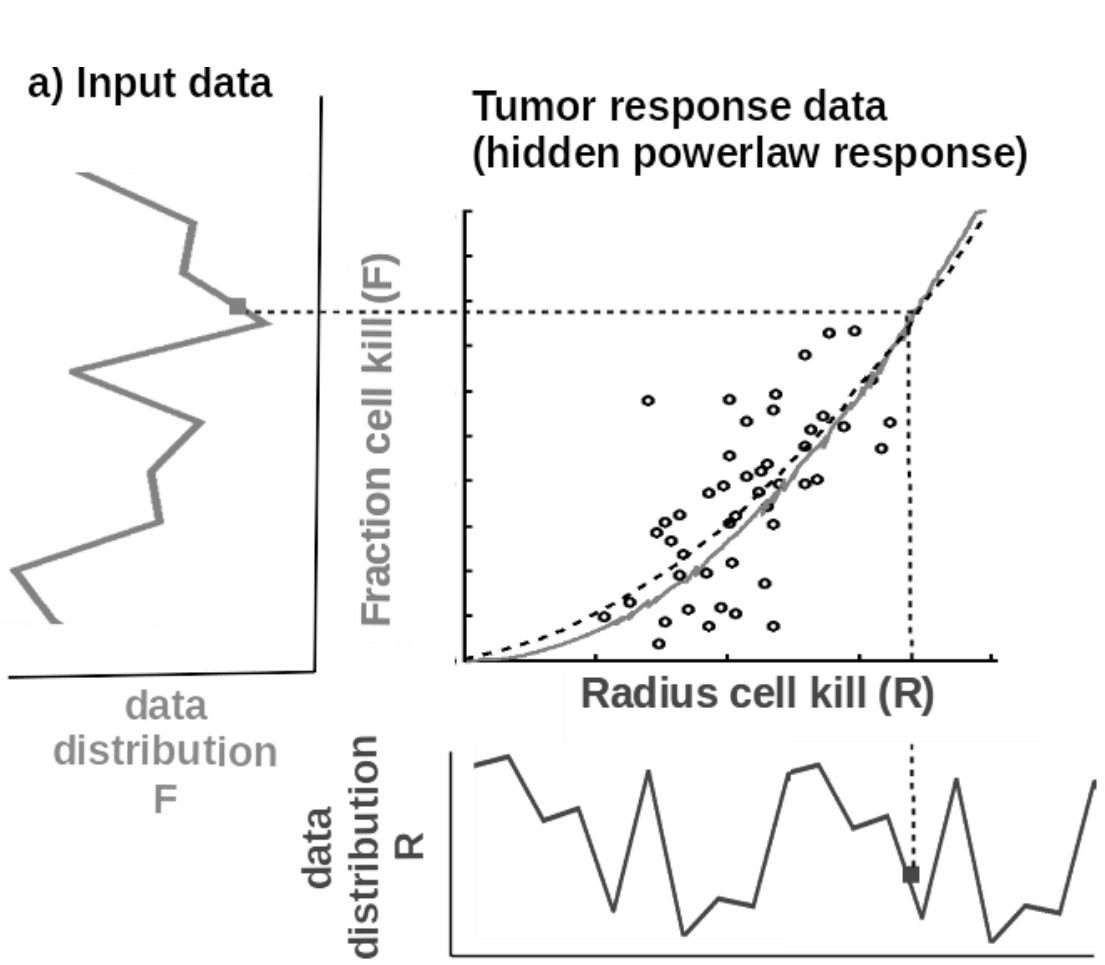
Figure: Treatment Algorithm for Stage I-III HER2-Positive Breast Cancer—AC = doxorubicin, cyclophosphamide; FEC = fluorouracil, epirubicin, cyclophosphamide; H = trastuzumab; P = paclitaxel; T = docetaxel; TCH = docetaxel, carboplatin, trastuzumab.

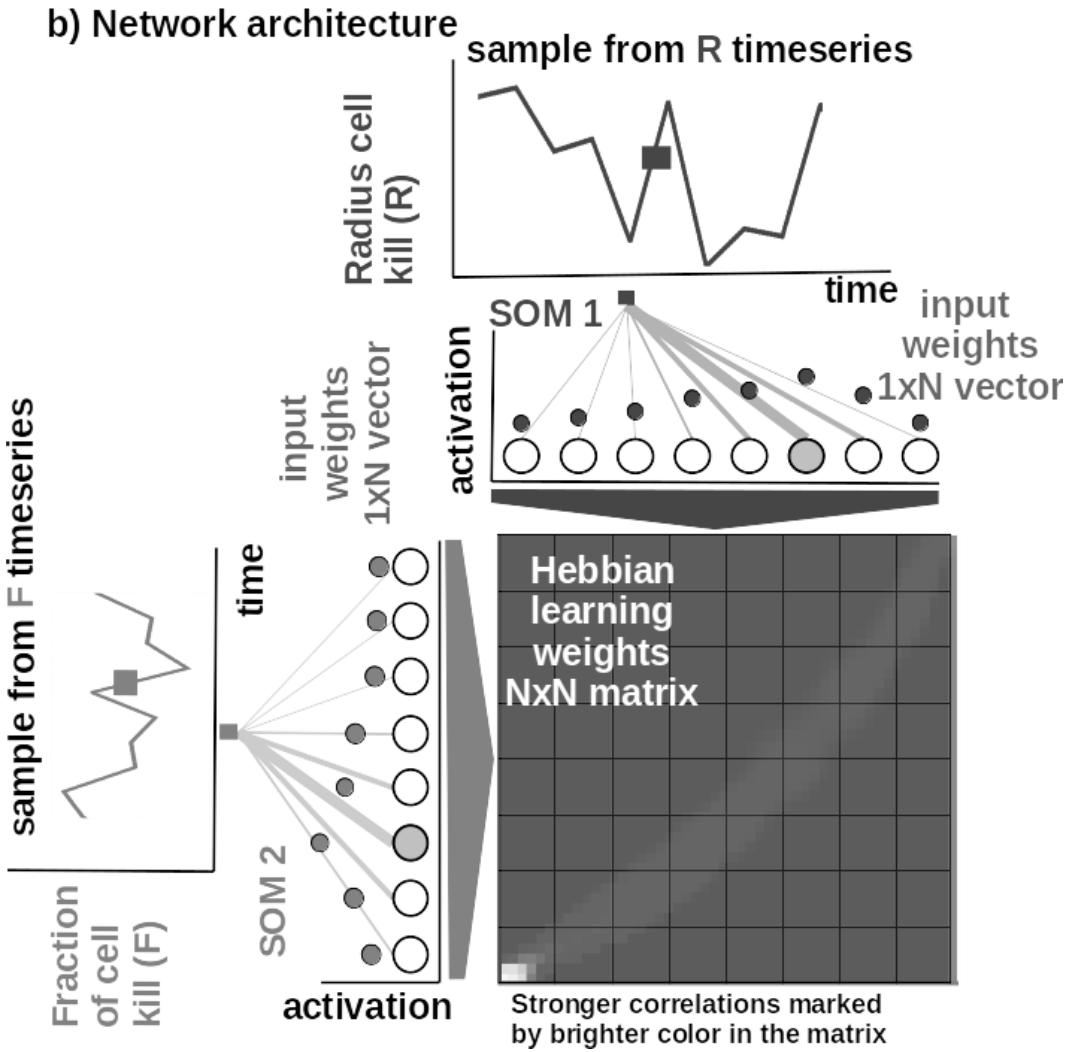
Tumor growth models

Growth under chemotherapy



Model instantiation



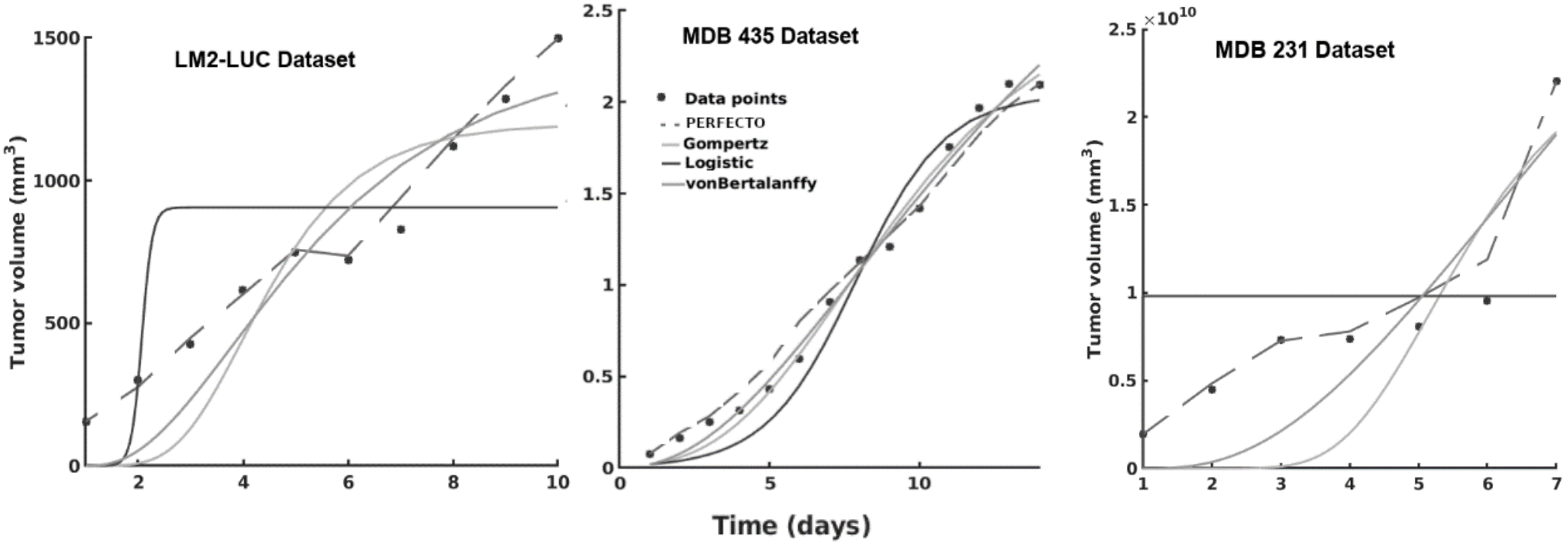


Experiments and evaluation

Learning unperturbed tumor growth

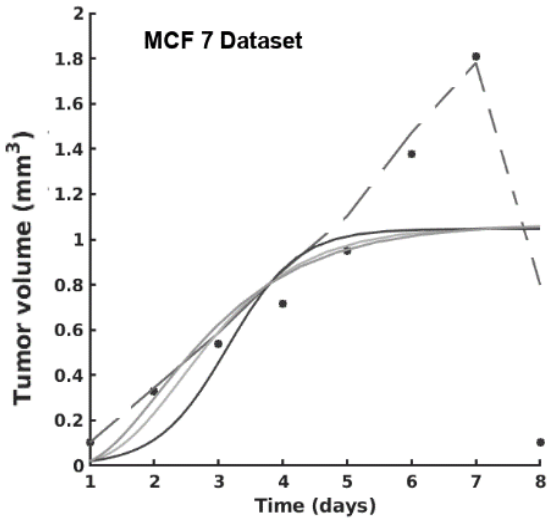
Experimental dataset setup

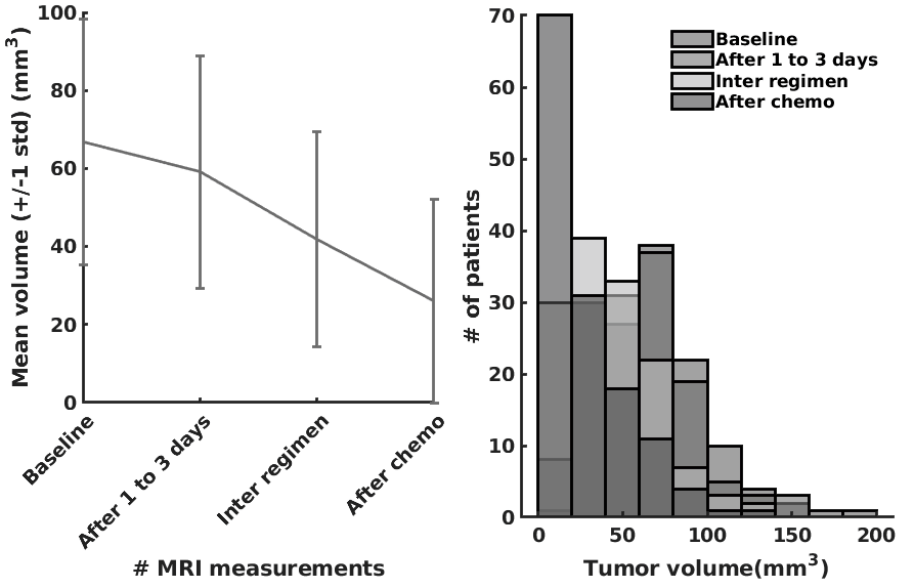
Dataset	Cancer Type	Data Type	Data Points	Data Freq.
1	MDA-MB-231 cell line	Fluorescence imaging	7	2x/week
2	MDA-MB-435 cell line	Digital Caliper	14	2x/week
3	MCF-7 cell line	Caliper	8	1x/week
4	LM2-4LUC+ cell line	Digital Caliper	10	3x/week



Experiments and evaluation

Learning perturbed tumor growth





Dataset/Model	SSE	RMSE	sMAPE
MDA-MB-231 cell line cancer [29]			
Logistic	7009.6	37.4423	1.7088
Bertalanffy	8004.9	44.7350	1.7088
Gompertz	7971.8	39.9294	1.7088
PERFECTO	119.3	4.1285	0.0768
MDA-MB-435 cell line cancer [15]			
Logistic	0.2936	0.1713	0.1437
Bertalanffy	0.2315	0.1604	0.1437
Gompertz	0.3175	0.1782	0.1437
PERFECTO	0.0977	0.0902	0.0763
MCF-7 cell line cancer [30]			
Logistic	3.0007	0.7072	1.0607
Bertalanffy	3.2943	0.8117	1.0607
Gompertz	3.1909	0.7293	1.0607
PERFECTO	0.7669	0.3096	0.2615
LM2-4LUC+ cell line cancer [31]			
Logistic	45.6032	2.3876	1.4816
Bertalanffy	56.0739	2.8303	1.4816
Gompertz	53.2428	2.5798	1.4816
PERFECTO	0.2009	0.1417	0.0365
I-SPY2 Trial [32]			
Logistic	248.3735	11.1439	1.7833
Bertalanffy	259.0963	16.0963	1.7834
Gompertz	260.3747	11.4100	1.7883
PERFECTO	0.8650	0.4650	0.0389

PRECISION

PROFILE EXTRACTION
FROM CLINICAL INSIGHTS
FOR SMART INDIVIDUALIZED ONCOLOGY

Data

309 Patients, 16 Clinical Features (https://simulacrum.healthdatainsight.org.uk/)

T_BEST	N_BEST	M_BEST	AGE	ER_SCORE	PR_SCORE	MORPH_ICD10_O2_DESC	GRADE_DESC	ER_STATUS_DESC	PR_STATUS_DESC	HER2_STATUS_DESC	HISTOLOGY_DESC
2	1	0	46.00	0	0	Intraductal carcinoma, noninfiltrating/Infiltr...	3 / III / POORLY DIFFERENTIATED	Negative	Negative	Positive	DUCT INFILTRATING CARCINOMA
2	0	0	83.00	8	8	Intraductal carcinoma, noninfiltrating/Infiltr...	2 / II / MODERATELY DIFFERENTIATED	Positive	Positive	Negative	DUCT INFILTRATING CARCINOMA
2	3a	1	79.00	8	6	Intraductal carcinoma, noninfiltrating/Infiltr...	2 / II / MODERATELY DIFFERENTIATED	Positive	Positive	Negative	DUCT INFILTRATING CARCINOMA
1c	0	0	75.00	8	8	Intraductal carcinoma, noninfiltrating/Infiltr...	2 / II / MODERATELY DIFFERENTIATED	Positive	Positive	Negative	DUCT INFILTRATING CARCINOMA
1c	0	0	85.00	8	8	Lobular carcinoma	2 / II / MODERATELY DIFFERENTIATED	Positive	Positive	Negative	CA LOBULAR NOS
...	...	...	...	...	...	...	...	...	...	...	...
1c	0	0	60.00	8	6	Lobular carcinoma	2 / II / MODERATELY DIFFERENTIATED	Positive	Positive	Negative	CA LOBULAR NOS
1b	0	0	68.00	8	8	Intraductal carcinoma, noninfiltrating/Infiltr...	3 / III / POORLY DIFFERENTIATED	Positive	Positive	Positive	DUCT INFILTRATING CARCINOMA
1b	0	0	75.00	0	0	Intraductal carcinoma, noninfiltrating/Infiltr...	3 / III / POORLY DIFFERENTIATED	Negative	Negative	Negative	DUCT INFILTRATING CARCINOMA
1c	0	0	65.00	8	7	Lobular carcinoma	2 / II / MODERATELY DIFFERENTIATED	Positive	Positive	Negative	CA LOBULAR NOS
1c	0	0	55.00	0	0	Intraductal carcinoma, noninfiltrating/Infiltr...	3 / III / POORLY DIFFERENTIATED	Negative	Negative	Negative	DUCT INFILTRATING CARCINOMA

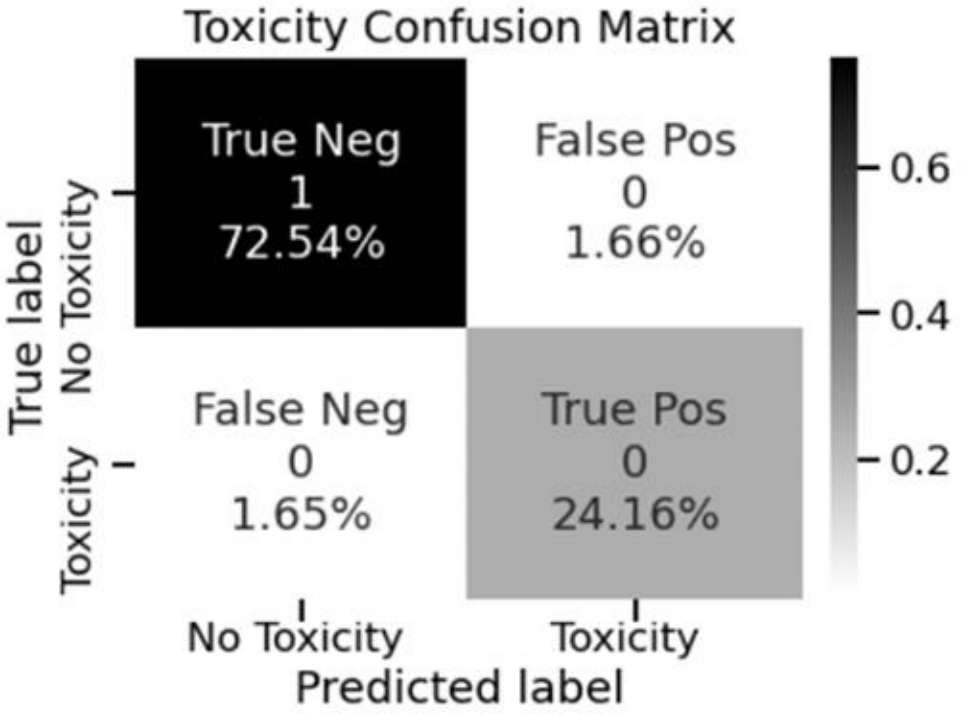
Data

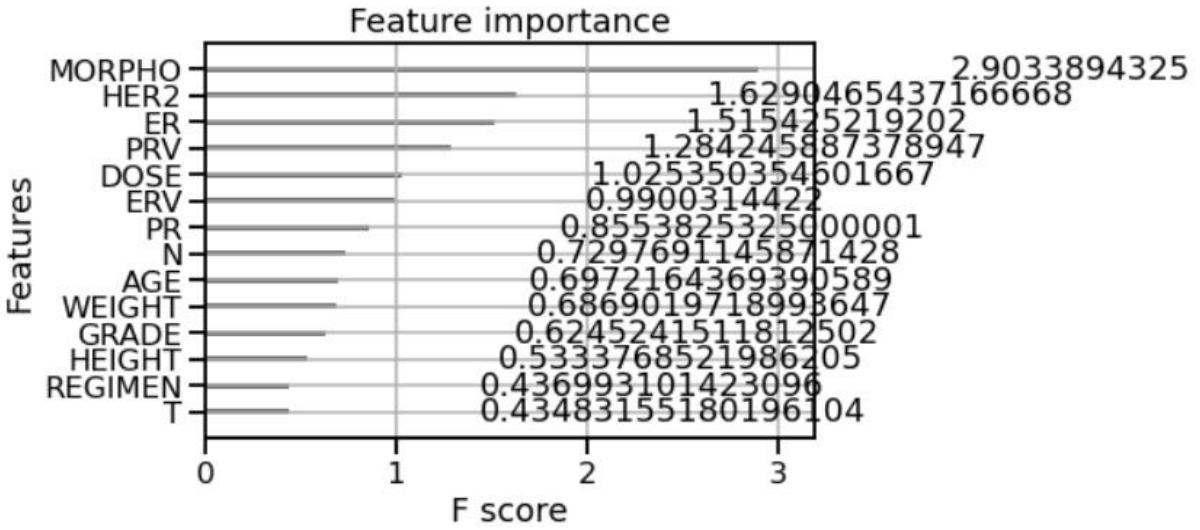
309 Patients, 16 Clinical Features

HEIGHT_AT_START_OF_REGIMEN	WEIGHT_AT_START_OF_REGIMEN	MAPPED_REGIMEN	REGIMEN_OUTCOME_SUMMARY	ACTUAL_DOSE_PER_ADMINISTRATION
1.65	77.60	Trial Unspecified	0	160.00
1.57	74.00	CMF	0	200.00
1.50	64.00	Trastuzumab Subcutaneous	0	135.00
1.55	67.40	FEC 100	0	120.00
1.72	79.00	Trastuzumab 21 day loading dose	0	1000.00
...	...	...	...	...
1.71	73.00	FEC + TRASTUZUMAB	0	1000.00
1.49	70.70	FEC 60 OR 75 + DOCETAXEL	0	142.00
1.70	58.00	Carboplatin + Paclitaxel (3 weekly)	0	800.00
1.62	119.00	CMF IV (28 day)	0	140.00
1.62	62.00	Carboplatin + Paclitaxel (3 weekly)	0	800.00

Model performance

Fine tuned ensemble model (Interacting Computational Maps + XGBoost)





Model insights

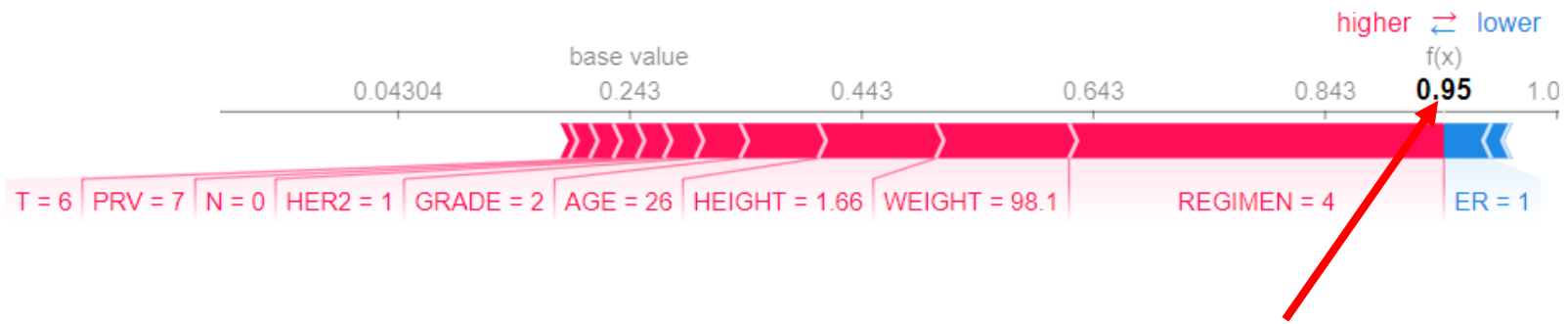
Sample **positive** patient data

T_BEST	N_BEST	M_BEST	AGE	ER_SCORE	PR_SCORE	MORPH_ICD10_O2_DESC	GRADE_DESC	ER_STATUS_DESC	PR_STATUS_DESC	HER2_STATUS_DESC	HISTOLOGY_DESC
2	0	0	26.00	8	8	Intraductal carcinoma, noninfiltrating/Infiltr...	3 / III / POORLY DIFFERENTIATED	Positive	Positive	Negative	DUCT INFILTRATING CARCINOMA

HEIGHT_AT_START_OF_REGIMEN	WEIGHT_AT_START_OF_REGIMEN	MAPPED_REGIMEN	REGIMEN_OUTCOME_SUMMARY	ACTUAL_DOSE_PER_ADMINISTRATION
1.66	98.10	CYCLOPHOSPHAMIDE + DOCETAXEL + EPIRUBICIN	1	1300.00

1 Toxicity
0 NO Toxicity

Impact of each **feature** in the **prediction**



Probability that the patient has Toxicity

Model insights

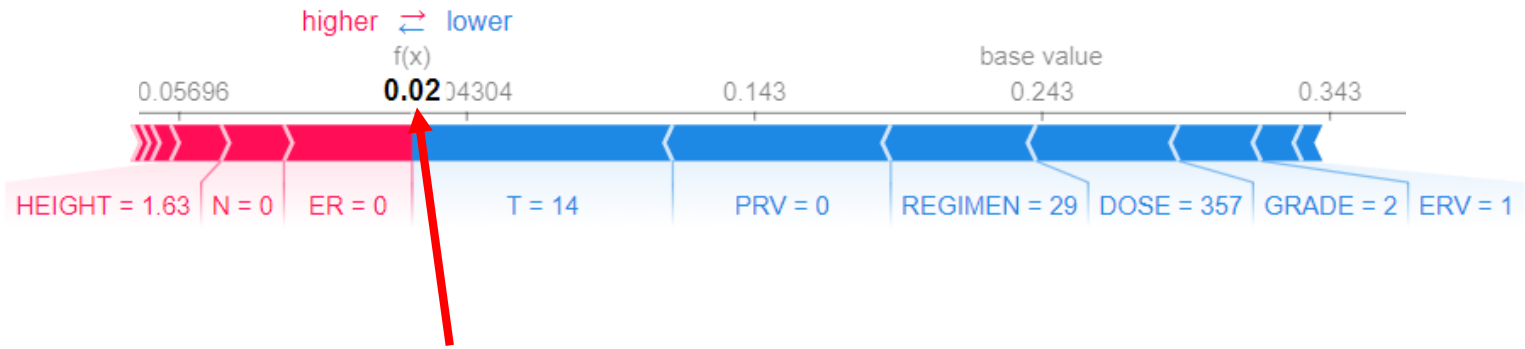
Sample **negative** patient data

T_BEST	N_BEST	M_BEST	AGE	ER_SCORE	PR_SCORE	MORPH_ICD10_O2_DESC	GRADE_DESC	ER_STATUS_DESC	PR_STATUS_DESC	HER2_STATUS_DESC	HISTOLOGY_DESC
1c	0	0	66.00	2	0	Intraductal carcinoma, noninfiltrating/Infiltr...	3 / III / POORLY DIFFERENTIATED	Negative	Negative	Negative	DUCT INFILTRATING CARCINOMA

HEIGHT_AT_START_OF_REGIMEN	WEIGHT_AT_START_OF_REGIMEN	MAPPED_REGIMEN	REGIMEN_OUTCOME_SUMMARY	ACTUAL_DOSE_PER_ADMINISTRATION
1.63	90.00	FEC 100	0	357.00

1 Toxicity
0 NO Toxicity

Impact of each **feature** in the **prediction**



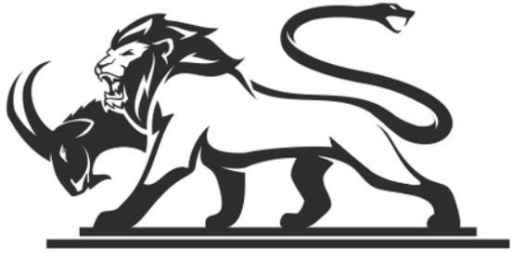
Probability that the patient has Toxicity

A Framework for Mathematical and Computational Oncology

Interacting Computational Maps

GLUECK

TUCANN



CHIMERA

PERFECTO

PRECISION

PROFILE EXTRACTION
FROM CLINICAL INSIGHTS
FOR SMART INDIVIDUALIZED ONCOLOGY