CHIC
Computational Horizons in Cancer

Torino, 9 February 2018

An RBF-based PSO approach for modeling prostate cancer

Ilaria Stura
PhD
ilaria.stura@unito.it

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No 600841

CHIC

Computational Horizons in Cancer

- The CHIC Project (2013-2017) was an **European Project** to create **mathematical models** on **cancer growth and response to treatment**

- **Joint work** between mathematicians, clinicians and biologist

- Models: Nephroblastoma, Glioblastoma, Lung and **Prostate cancer**s

Prostate cancer

Prostate cancer is a very common disease in man.

Fortunately, it:

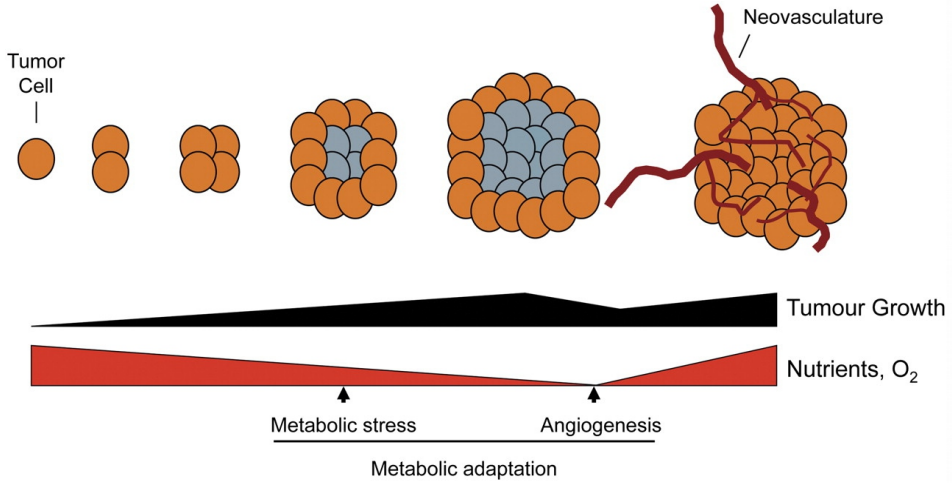
- Grows slowly
- Can be early diagnosed
- Can be treated with radiotherapy or surgery, eventually with hormone therapy
- In case of surgery, the relapse could be monitored by PSA exam

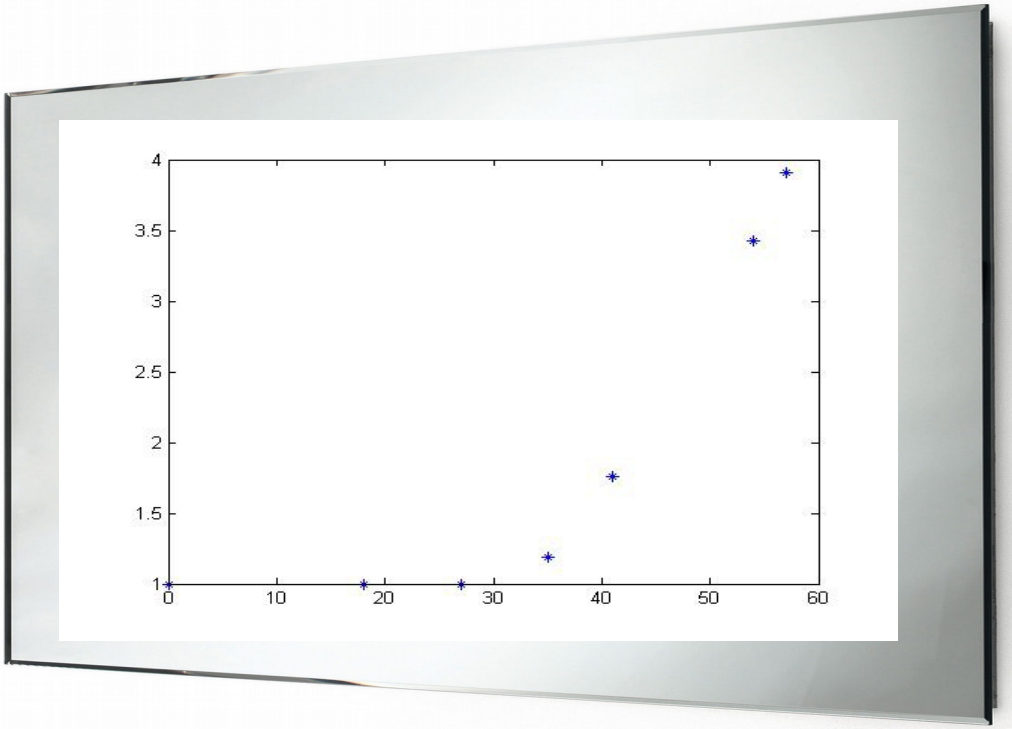
Available data

EUREKA1 study, made by our group, collects a large number of data of prostatectomized human patients.
In particular:

- Stage of the tumor
- Gleason Score (hystologic scoring of the removed tumor)
- Risk parameters (positive margins, lymph nodes…)
- Adjuvant therapies
- PSA: each 4-6 months patients should make the Prostate Specific Antigen exam; it is a good biomarker for the relapse

The usage of PSA





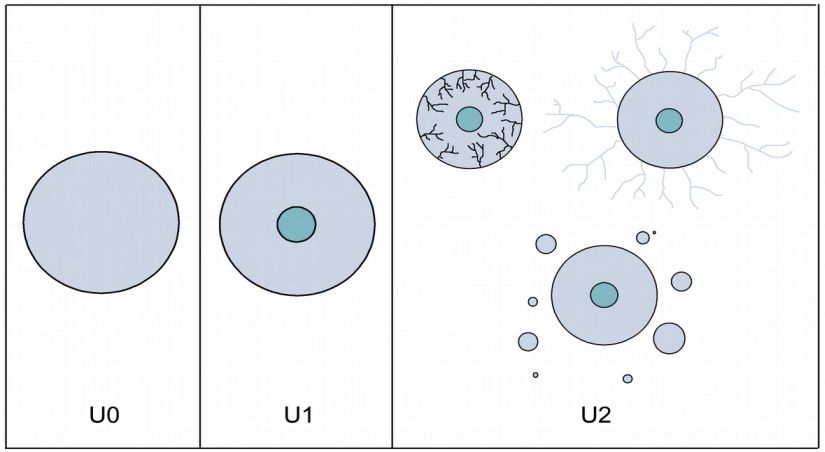
We can use, in case of prostate cancer, the PSA as a MIRROR of the real volume of the tumor.

But pay attention... it is not a perfect mirror... (we can't see the 'vampire'!)

Mathematical model

In our Phenomenological Universalities (PUN) approach we consider the following general law:

$$\begin{cases} \dfrac{dN(t)}{dt} = c(t)N(t) \\ \dfrac{dc(t)}{dt} = \displaystyle\sum_{i=0}^{n} \beta_i c^i \end{cases}$$



In which N is the size of the tumor and c(t) is the function of the growth coefficient, which derivative is described as a Taylor expansion of c.
When n=0, N is the exponential growth law; when n=1 N is the Gompertzian growth law and n=2 is the West growth law.

The mathematical model

Different possible grow functions: Malthus, Gompertz and West.

Now we focus on Gompertz:

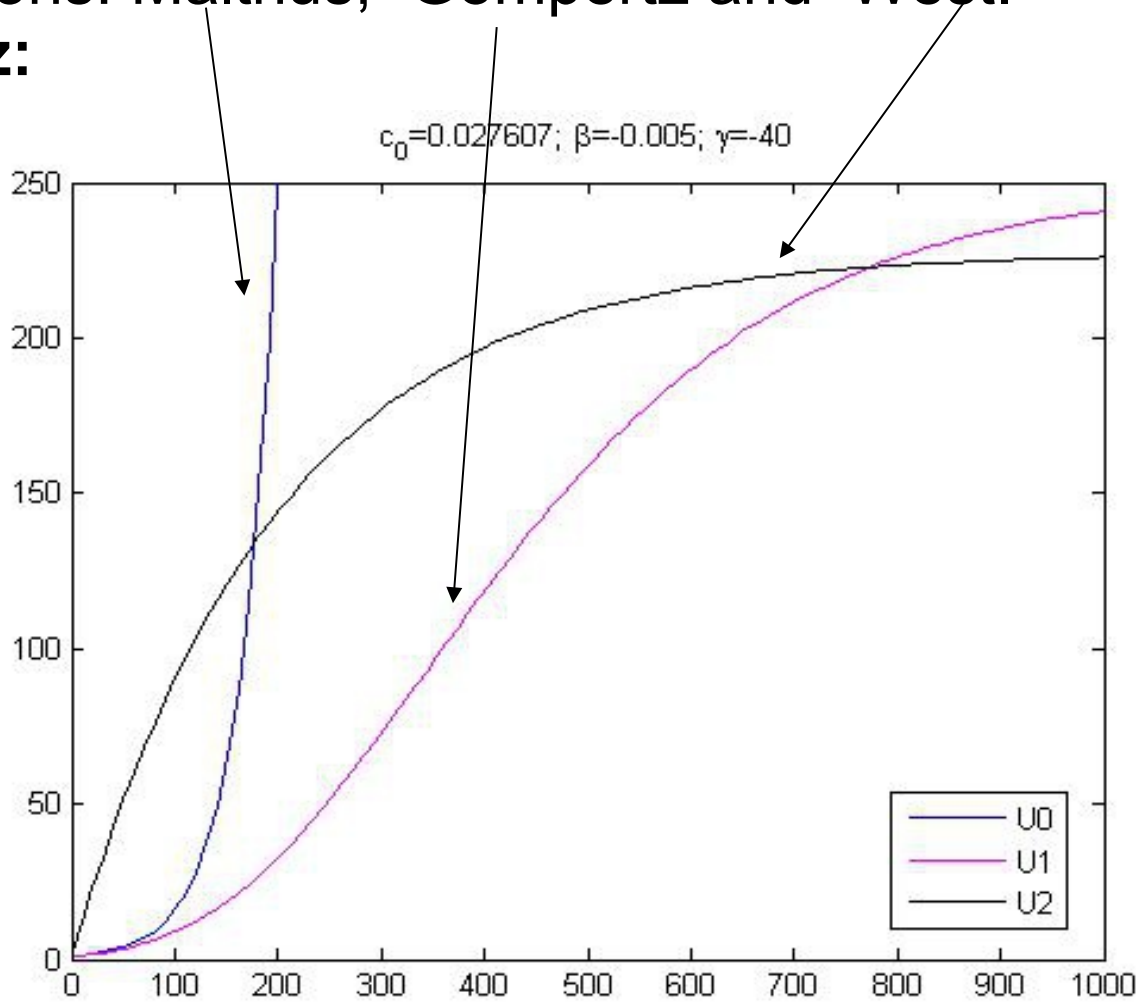
$$\begin{cases} \dfrac{dN(t)}{dt} = c(t)N(t) \\ \dfrac{dc(t)}{dt} = \beta c \end{cases}$$

Or:

$$\frac{dN(t)}{dt} = c_0 e^{\beta t} N(t)$$

With solution:

$$N(t) = N_0 e^{\frac{c_0}{\beta}(e^{\beta t}-1)}$$



Questions

- Is a Gompertz function a good approximation of our data?
- How we can estimate the growth parameters?
- Patients with similar clinical characteristics have similar growth parameters?
- Are the growth parameters predictive factors of a relapse after the therapy?

Questions

- Is a Gompertz function a good approximation of our data?
- How we can estimate the growth parameters?

- Patients with similar clinical characteristics have similar growth parameters?
- Are the growth parameters predictive factors of a relapse after the therapy?

Particle Swarm Optimization

It is a cooperative method first introduced by Kennedy (social psychologist) and Eberhart (electrical engineer).

Using a stochastic method we can simulate trajectories of all singles birds by considering their **selfish** (to explore by own, to remember where food was more abundant) and **social** (to return in the swarm and share information) behaviour and consequently to simulate the motion of the swarm.

If a good trade-off between the two behaviours is allowed then the flock can reach the minimum (i.e. the place with maximum availability of food).

Swarm Intelligence

1. We initialize randomly the position p_i of the bees in the search space with a random velocity v_i
2. At each step, we update velocity and position:

$$v_i^{(j)} = \omega^{(j)} v_i^{(j-1)} + \varphi_l^{(j)} (l_i^{(j-1)} - p_i^{(j-1)}) + \varphi_g^{(j)} (g_i^{(j-1)} - p_i^{(j-1)}),$$

$$p_i^{(j)} = \omega p_i^{(j-1)} + v_i^{(j)}.$$

Where j is the number of iterations, g is the global best position, l is the best position of the single bee, ω is the inertia weight and ϕ_l, ϕ_g are the cognitive and social behaviours.

Swarm Intelligence

In our case, the objective function is:

$$\sum_{i=1}^{n} \left[N_i - N_0 e^{\frac{c_0}{\beta}(e^{\beta t_i} - 1)} \right]^2,$$

And we want to find the best values of c_0 and β.

Problem: this method needs a large cardinality of the sample (n) but → the number of available PSA is 6 (in average).

We create a larger sample reconstructing the PSA curve with RBF

RBF

- Given a set $\mathcal{X}_N = \{(x_i, y_i) \in \mathbb{R}^2, i = 1, \ldots, N\}$ of N distinct *nodes*, in a domain $\Omega \subseteq \mathbb{R}^2$, and a set $\mathcal{F}_N = \{f_i = f(x_i, y_i), i = 1, \ldots, N\}$ of *data values*, the standard RBF interpolation problem consists in finding an interpolant $R : \Omega \to \mathbb{R}$ of the form:

$$R(x, y) = \sum_{i=1}^{N} c_i \phi(||(x, y) - (x_i, y_i)||_2), \quad (x, y) \in \Omega, \tag{1}$$

  where $|| \cdot ||_2$ is the Euclidean norm, and $\phi : [0, \infty) \longrightarrow \mathbb{R}$ is a RBF.

- The coefficients $\{c_i\}_{i=1}^{N}$ are determined by enforcing the interpolation conditions $R(x_i, y_i) = f_i$, $i = 1, \ldots, N$.

- This leads to a symmetric linear system of equations:

$$\Phi \boldsymbol{c} = \boldsymbol{f}, \tag{2}$$

  where $\Phi_{ki} = \phi(||(x_k, y_k) - (x_i, y_i)||_2)$, $k, i = 1, \ldots, N$.

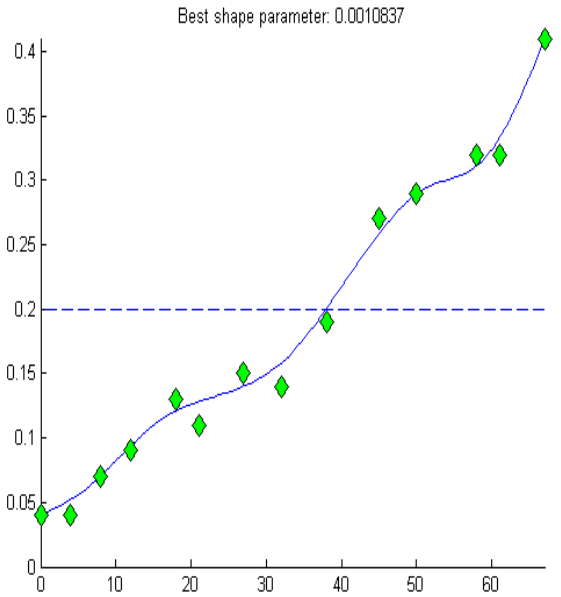
RBF – PSO

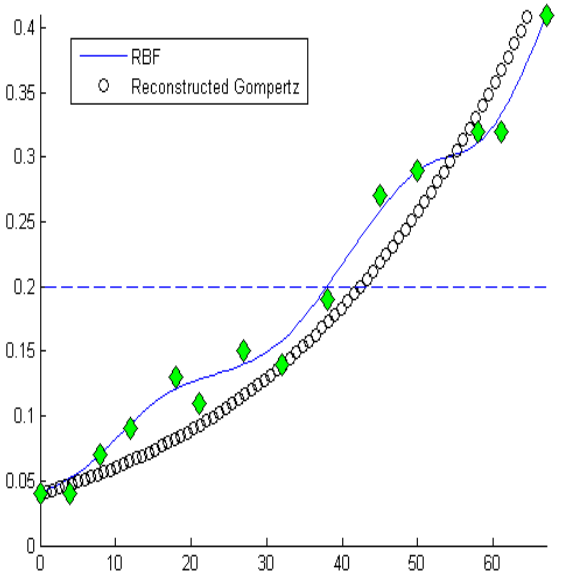
- Radial Basis Functions (RBFs) to have more data

  Robust: OK also in case of non-monotonic and irregularly spaced data!

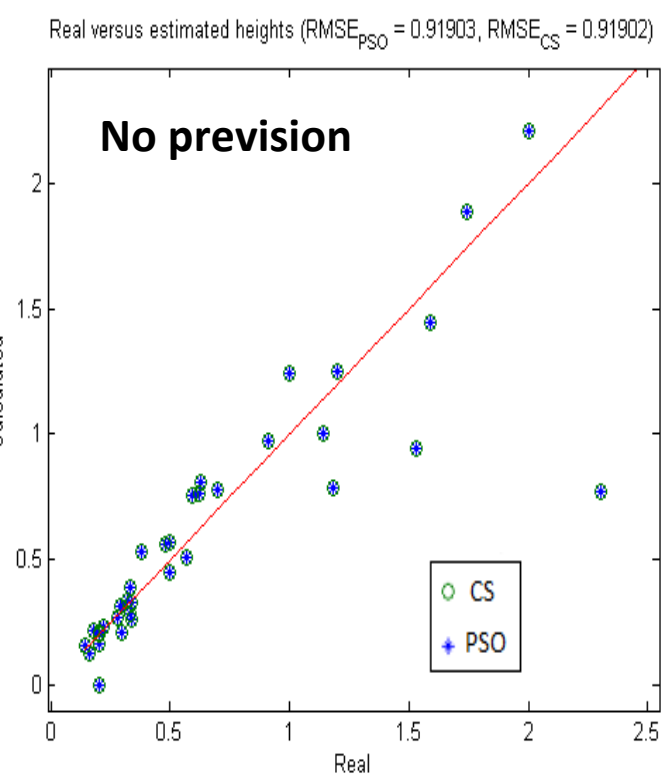
- Particle Swarm Optimization (PSO) to estimate the parameters

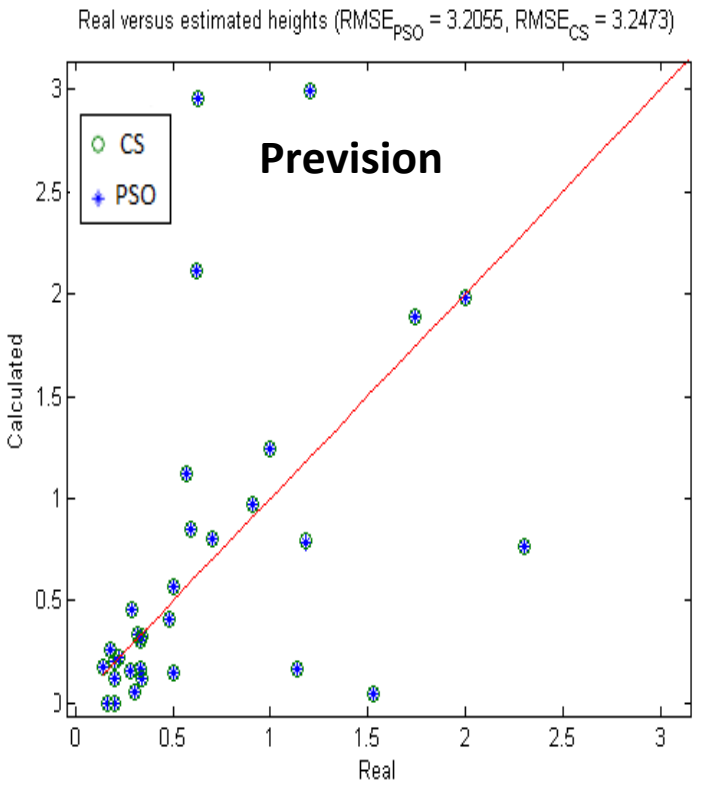
  (Gompertzian) curve reconstruction





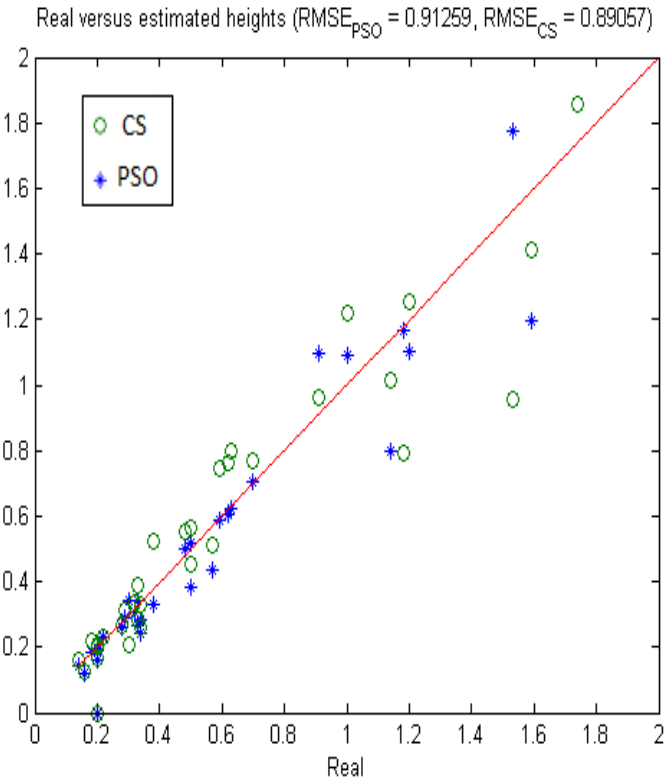
Accuracy



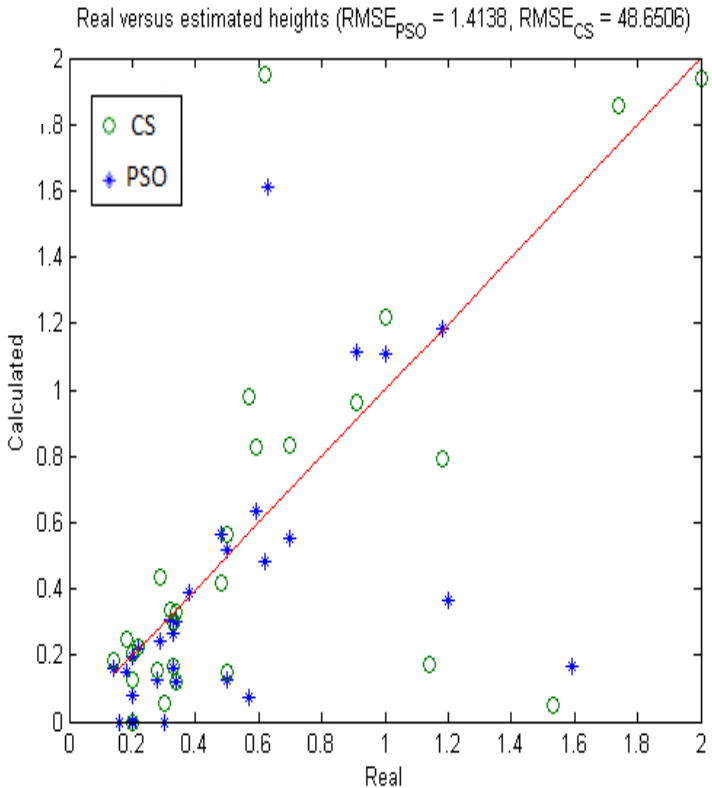


1 parameter

$$\mathrm{RMSE} = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$



2 parameters



Data set

pathological Gleason Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	6	1.49	6	1.49
4	3	0.74	9	2.23
5	27	6.68	36	8.91
6	82	20.30	118	29.21
7	179	44.31	297	73.51
8	66	16.34	363	89.85
9	40	9.90	403	99.75
10	1	0.25	404	100.00

pathological stage	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NA	35	8.66	35	8.66
pT2a	44	10.89	79	19.55
pT2b	18	4.46	97	24.01
pT3c	148	36.63	245	60.64
pT3a	79	19.55	324	80.20
pT3b	77	19.06	401	99.26
pT4	3	0.74	404	100.00

404 patients at the beginning...

Adjuvant HT (months)	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No HT	313	77.48	313	77.48
< 6 months of HT	6	1.49	319	78.96
6 < months of HT <= 12	10	2.48	329	81.44
12 < months of HT <= 24	24	5.94	353	87.38
> 24 months of HT	51	12.62	404	100.00

Adjuvant RT	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	326	80.69	326	80.69
Yes	78	19.31	404	100.00

True time to relapse	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	47	11.63	47	11.63
T <= 24 months	160	39.60	207	51.24
24 < T <= 48 m	121	29.95	328	81.19
48 < T <= 72 m	47	11.63	375	92.82
T > 72 months	29	7.18	404	100.00

Data set

pathological Gleason Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	4	1.61	4	1.61
4	3	1.21	7	2.82
5	22	8.87	29	11.69
6	58	23.39	87	35.08
7	126	50.81	213	85.89
8	24	9.68	237	95.56
9	11	4.44	248	100.00

True time to relapse	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	14	5.65	14	5.65
T <= 24 months	117	47.18	131	52.82
24 < T <= 48 m	64	25.81	195	78.63
48 < T <= 72 m	30	12.10	225	90.73
T > 72 months	23	9.27	248	100.00

pathological stage	Frequency	Percent	Cumulative Frequency	Cumulative Percent
NA	29	11.69	29	11.69
pT2a	37	14.92	66	26.61
pT2b	14	5.65	80	32.26
pT3c	112	45.16	192	77.42
pT3a	48	19.35	240	96.77
pT3b	8	3.23	248	100.00

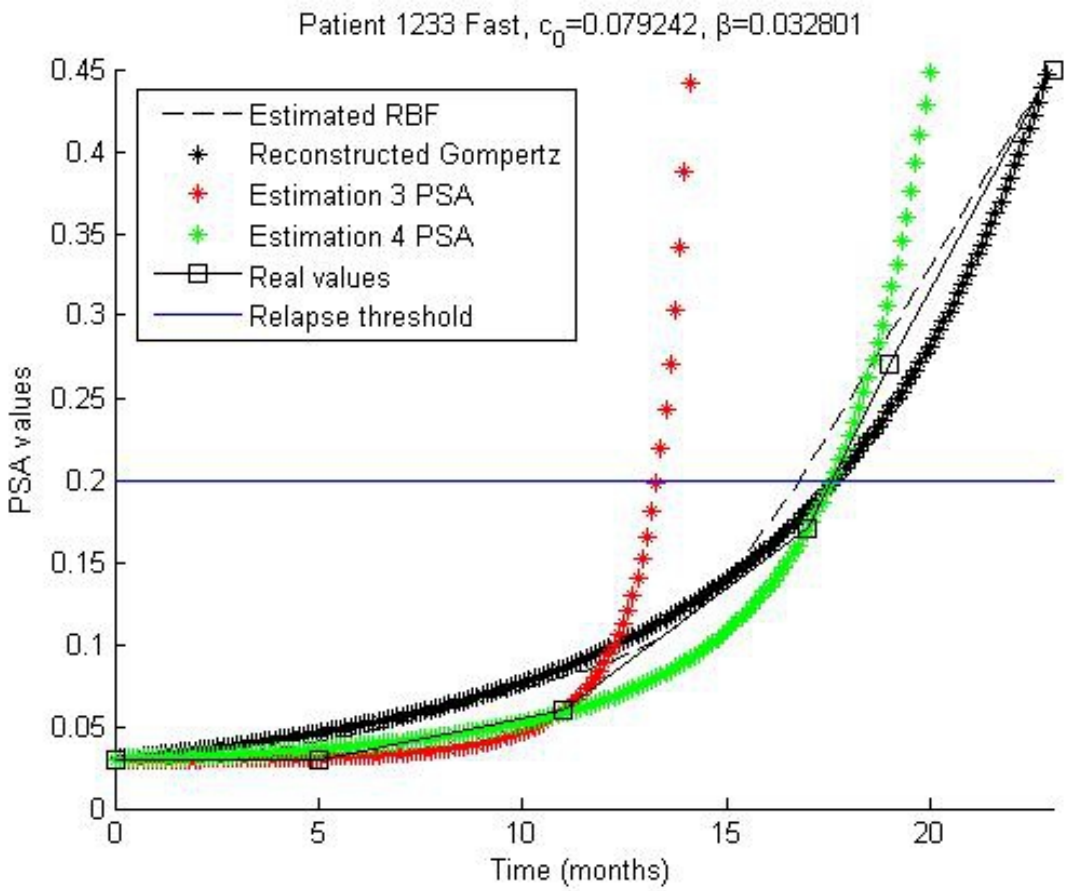
248 patients without adjuvant therapies and with a successful surgery (first PSA after surgery < 0.2)

Results

Using the complete series, we successfully reconstruct the real PSA series with the estimated parameters.

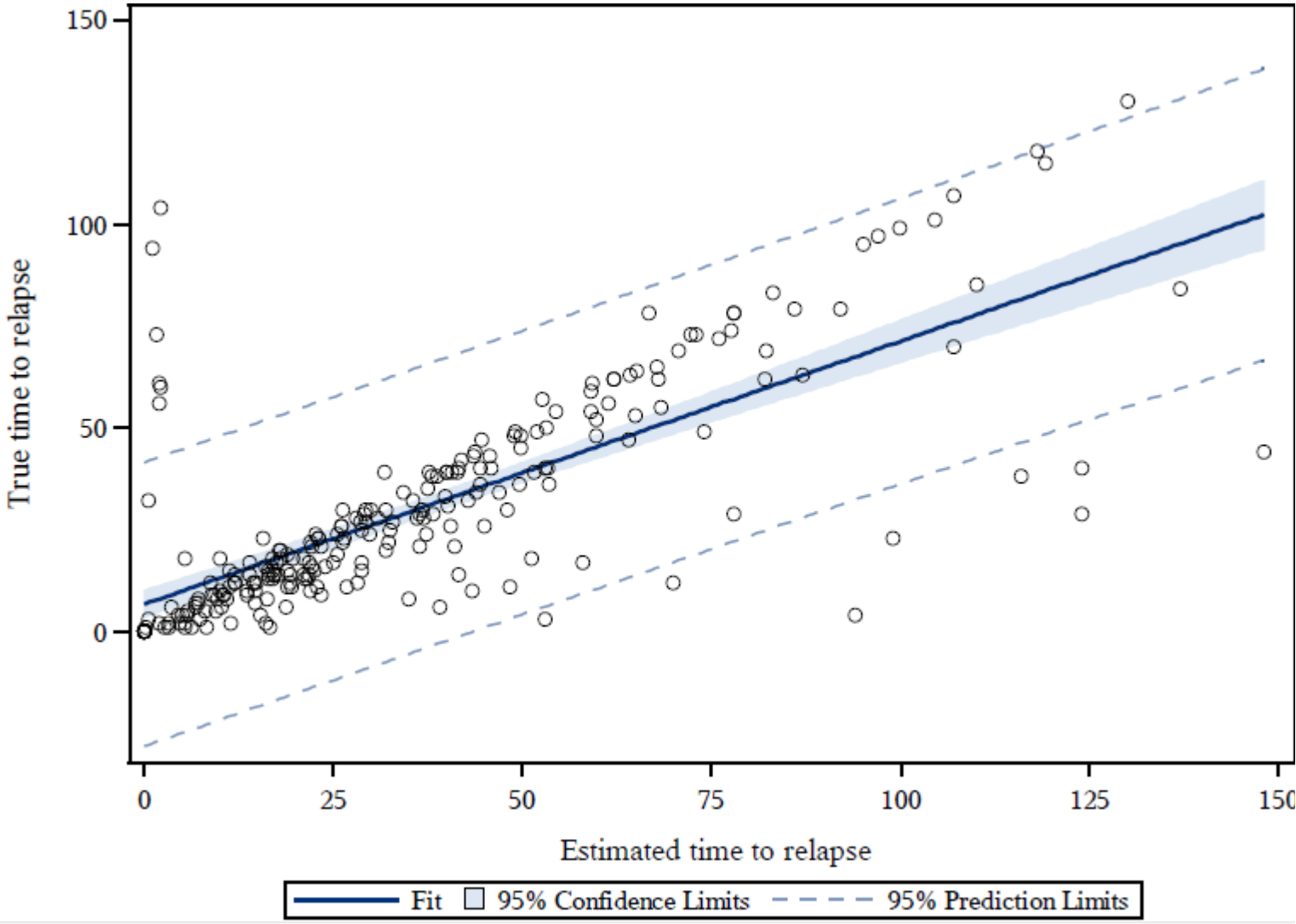
Using only 3 PSA values, the estimation is not accurate.

Using 4 PSA values, in the majority of cases we can predict the time to relapse of the patient.



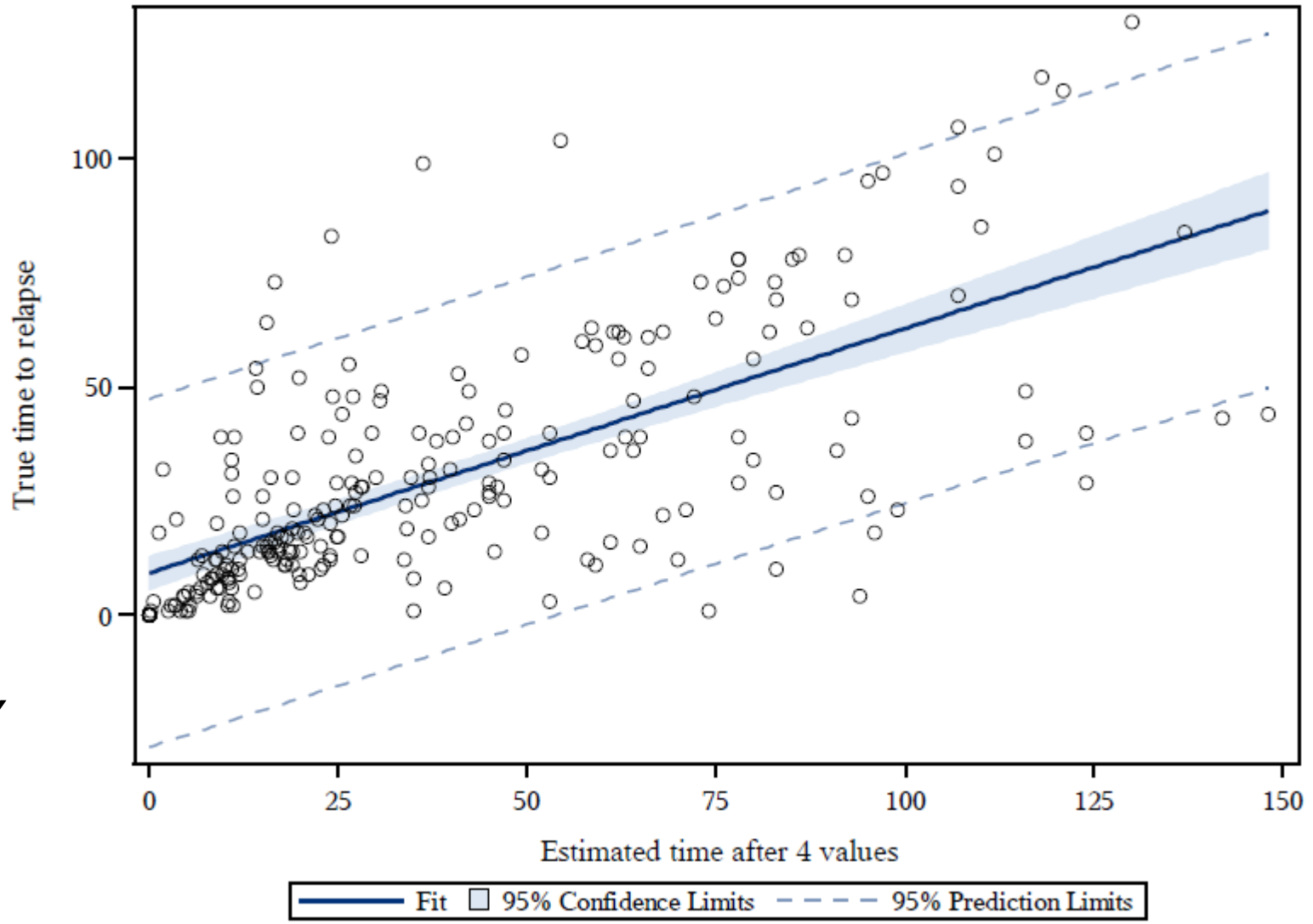
Results

High correlation between the real time to relapse (indicated by clinicians) and the estimated time to relapse (curve created using estimated c0 and beta)

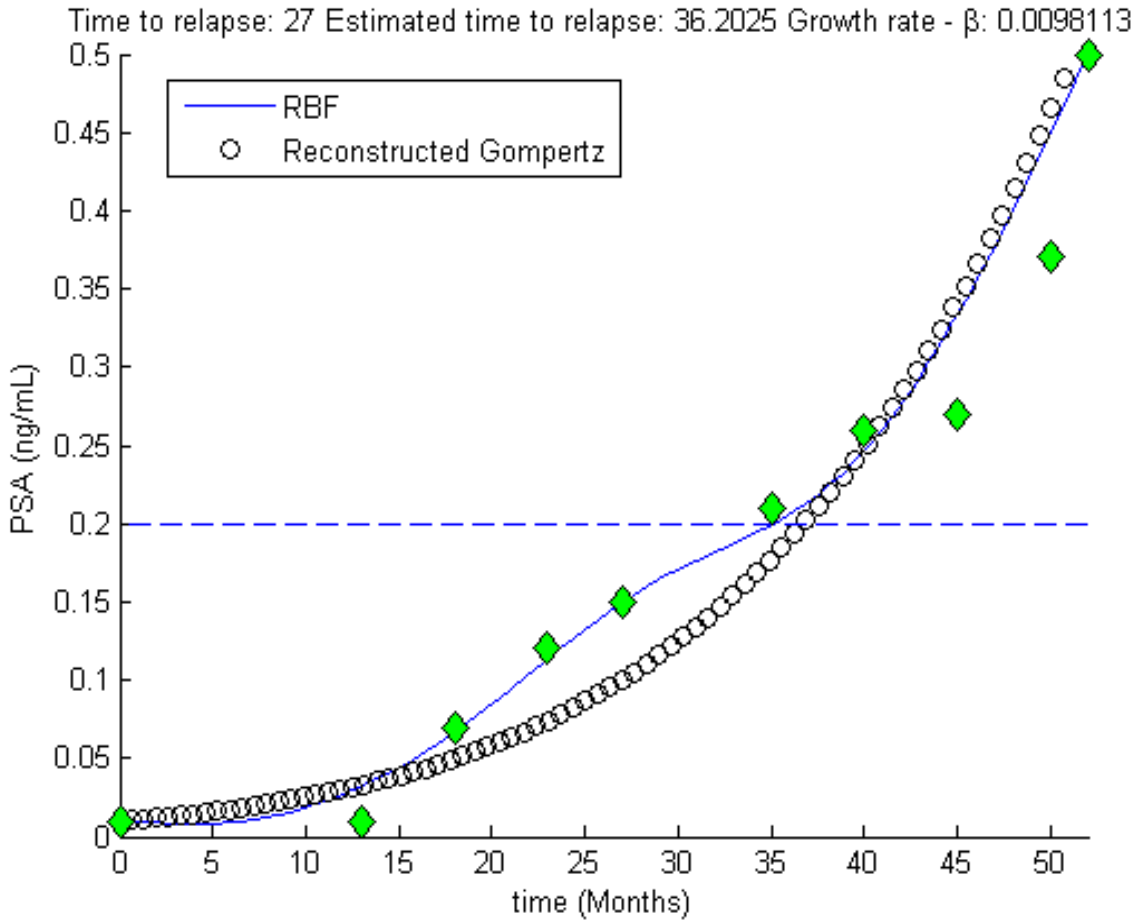


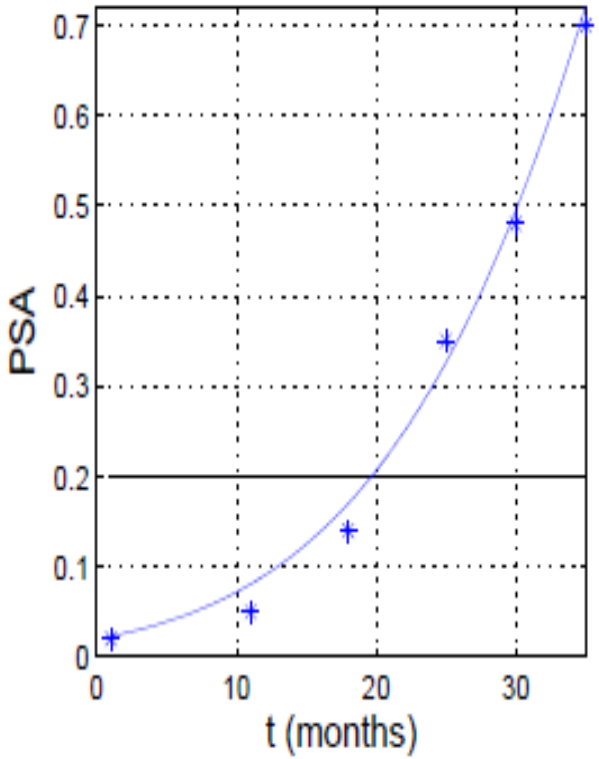
Results

High correlation between the real time to relapse (indicated by clinicians) and the estimated time to relapse (curve created using estimated c0 and beta AFTER ONLY 4 VALUES)



Examples of output



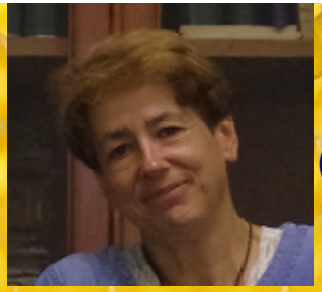


Conclusions

- PSO can be used to find the parameters of a Gompertzian function describing the growth of PSA values.
- RBF can be used to increase the size of the sample.
- Using only 4 PSA values we can have a prediction of the timing of relapse of the patient.

- We must pay attention to the series: duplicate, too distant in time, oscillating values could affect the estimation of the parameters!!!

Thank you for your attention!



Caterina Guiot (the boss, physicist)

Ezio Venturino (Professor, mathematician)

Ilaria Stura (PhD, mathematician)

Domenico Gabriele (PhD, clinician)

A bee of the swarm (mascot, worker)

Emma Perracchione (PhD, mathematician)