TIME BEHAVIOUR OF PROSTATA TUMOR MARKERS: FROM MODELLING TO A WEB TOOL FOR PHYSICIANS

Nikolas Popper[1,2], Günther Zauner[1,2], Veronika Bodmann[1]

[1]Vienna University of Technology, Institute for Analysis and Scientific Computing,
1040 Vienna, Wiedner Hauptstraße 8-10 , Austria
[2]"die Drahtwarenhandlung", Neustiftgasse 57-59, 1070 Vienna, Austria

niki.popper@drahtwarenhandlung.at (Nikolas Popper)

Abstract

This paper deals with mathematical optimization and modelling in the area of medical examination for prostate cancer and how to develop a web application to help physicians in case of diagnosis and for definition the next appointment for an appraisal.

The PSA marker can be measured via a simple blood test and is considered to have exponential behaviour from the fitting function point of view. But this behaviour is assumed to be exponential, when a person has prostate cancer or in the case that after a prostate ablation a part of the tumor is left in the body of the patient. The doubling time of this marker should be calculated out of free starting values and their x – axis value in months. For the whole system a PHP – framework is provided and a connection from this so defined graphical input/output – interface to a MATLAB Webserver is done. The first task is a nonlinear fit problem. We have to find parameters *a, b* and *c* as real numbers, so that the function

$$f(time) = a + b * exp(c * time)$$

has a good least square error fitting to the measurements. Our solution is the approximation for the doubling time of the value in the last measurement point. This can be done easily by a numerical fixed point iteration or for the redefined model as a zero crossing detection. As an extension the effects of measurements errors are shown by calculating the Gaussian – Variance – Analysis. In the summary other possibilities for using the implemented interface to support physicians with web tools for simulation are presented.

Keywords: PSA, web application, nonlin. fit, PHP/MATLAB interface, prostate cancer

Presenting Author's biography

Nikolas Popper. He has earned a degree in technical mathematics at the Vienna Univerity of Technology. He has experience in industry projects as well as research and development knowledge. Currently he is working in the area of visualization in computer graphics and modeling and simulation of epidemics. He is co-proprietor of the company 'die Drahtwarenhandlung' and offers as well technical solutions (defect detection on pictures, simulations, …) as animations and films. Furthermore he is doing a PhD thesis in the area of alternative and coupled models.

1 Medical Background

The prostate-specific antigen (PSA) is a protein produced by the prostate gland. The blood concentration of PSA can easily be measured by a simple blood test.

Usually the concentration of PSA is very low and is assumed to increase linearly when men get older. However, as also the tumors of prostate cancer and their metastasis produce PSA, so the PSA level in the blood can be taken as tumor marker. In these cases the growing of the PSA is assumed to be exponential. As prostate cancer is one of the main cancer diseases for men, the focus on alternative or additional examinations is quite high.

Though there are also cases of prostate cancer with low PSA level, the risk is increasing considerably with the concentration. One important value is the so called doubling time, which shows how long it takes until the PSA level of a certain patient has doubled. With this time, the physician decides how long you can wait until the next PSA screening should be done.

The original challenge was to build a web-application for physicians, which allows to compute the doubling time for a patient with 2 measured PSA levels at 2 different times.

Additionally the question of accuracy arises so the results can be rated. From the mathematical point of view, we assumed that the results with only 2 measurements will not be very useful, as inaccuracy raises, or rather the used statistic formulas cannot be applied for such small values of measurements. For these reasons only values with $n > 2$, in the standard case of $n = 3$ were considered..

The web application for physicians should compute the doubling time out of these – at least - three values. It should help the physicians to define a date for the next controlling date and gives additional information about the health status of the patient.

2 Tasks

The assumption of the PSA value to be an exponential function was chosen after evaluation of the available literature [1]. Decisively in the practical application a rather too quick doubling time (by fitting of the exponential function) is to prefer to a too slow one. In this case - an application in the health area - the error likelihood must be in any case on the “safe” side, an unavoidable condition for the successful application of the tool.

So the first task is a nonlinear fit problem. We have to find parameters *a*, *b* and *c* as real numbers, so that the function Eq.(1)

$$f(time) = a + b * \exp(c * time) \quad (1)$$

has a good least square error fitting to the measurements. In our case it is of interest to find an approximation for the doubling time of the value in the last measurement point. After getting the parameters *a*, *b* and *c* this can be done easily by a numerical fixed point iteration or for the redefined model as a zero crossing detection.

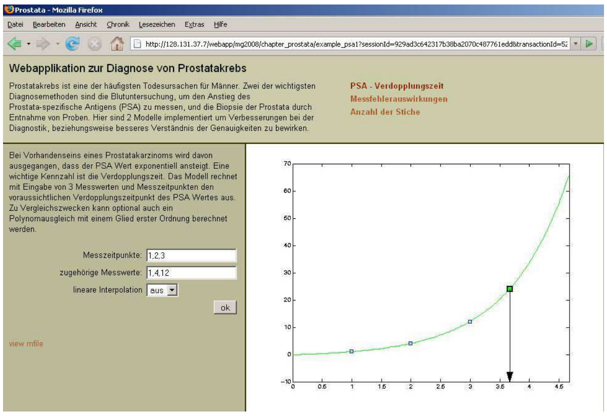

Fig. 1 Input of three values, computing doubling time with graphical output in a PHP – frame.

The best exponential fitting curve, according to the least square error measure, is computed through three measurement points. This is done to see when the doubling time of the last measurement will occur. As we can see in [Fig. 1] the result is plotted with a black arrow which depicts the calculated doubling time. To improve the convenience of using such a toll an internet based tool for assistance in analysis of the behaviour was developed for the users. Important additional features of the new tool, comparing with other systems developed in the USA [2], are the graphical output of the results for a better comprehension by the patients and the comparison with a linear approximation curve.

The graphical user interface is realized in PHP. It has a modular structure, which can easily be enhanced by additional examples. The parameter type tests of the input structure are done automatically. Moreover boundaries for every input variable can be defined, so that these additional work have not to be done for every application in the coupled computer numeric tool.

As a matter of fact the reliability of the results can be questioned as always for models in the area of health services. The first question is the principle of using an exponential fitted curve. Using this is a widely accepted approach, but still a matter of discussion. [3] As we describe later on the principle of using different functions for describing a physiological process (linear approach, exponential approach) defining it and discussing this problem with the user is one of the most important parts of this model. The second important question for using the tool is to sensitise physicians for the reliability of the results concerning the error estimation.

3 Error Estimation

The problem of a relatively high inaccuracy, caused by limited number of parameters was discussed with the users. Then the model was implemented in the web application, so that only the least number of the points - for simulation technical-mathematical reasons – was defined as three. The number of the points improves of course the reliability of the calculations.

As an extension of the model and to show the effects of measurements errors to the physicians, a corridor of error variation was implemented. This can be done by calculating the Gaussian – Variance – Analysis or, as done in this case by randomly displacing of the real measurements by an error afflicted signal. This is done as often as the user wants, by default 50 runs, and then the solutions corridor is plotted. The physician gets the interval of the anticipated doubling time back.

Furthermore this application can be used to show the physicians the possible effects on small measurement values and how rounding of small values can influence the results formidable. In addition for other implementations this approach can be used to show the importance of a minimum number of simulation runs. The user should learn about the effects occurring when the number of cycles for calculating the system with afflicted input signals is to low. For example, only a few experiments are calculated and in every run the measurements are changed to the lower direction. This can lead to an longer time span for the doubling time, then it is in reality.

Because the web solution's goal - beside the application in the teaching area – was the use in health care, different professional languages can lead to misunderstandings. The reliability of the data is the main goal, i.e. how exactly the result reflects the expectations of the treating doctor. Especially in the medical area results against intuition are hardly to argue.

To show the effect of the measurement errors, the variance is calculated with the Gaussian – Variance – Analysis: Eq.(2)

$$V(f(t)) = V(x_1, x_2, x_3, t) = \sqrt{v_1^{\,2} \cdot f_{x_1}^{\;\;2} + v_2^{\,2} \cdot f_{x_2}^{\;\;2} + v_3^{\,2} \cdot f_{x_3}^{\;\;2}} \quad (2)$$

However, to avoid time consuming calculations for the web-application this was implemented by randomly displacing of the real measurements by an error signal. This is done as often as the user wants, by default 50 runs, and then the solution corridor is plotted (see Fig. 2). The physician gets the interval of the anticipated doubling time back.

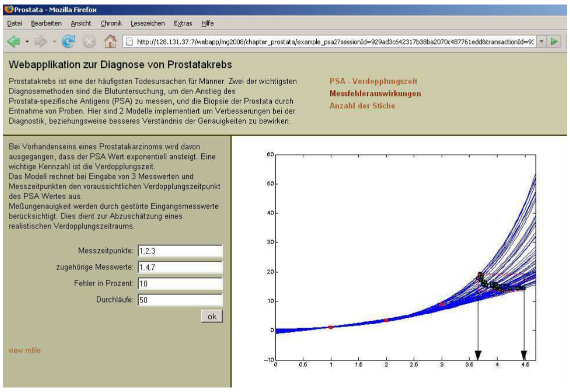

Fig. 2 Solutions with error afflicted input data, the two arrows in the right section of the figure, point out the minimal and maximal doubling time, as calculated for the exponential growth model assumption.

One of the main tasks and improvements to other methods is to question the reliability of the results.

The result, which is calculated through solving the solution routine with different randomly disturbed measurement data streams, is visualized by plotting all exponential interpolation curves and an interval, whereby the left boundary is the minimal doubling time and the right sides is the maximum doubling time. The length of the interval depends not only on the measurement inputs, but also on the allowed error in percent for every measurement point. As can be seen, especially for very small input data (for example after a carcinoma ablation) measurement errors lead to long intervals. This means that the PSA marker measurement is only an adequate method for such patients, if the physician can guarantee low measuring faults. This is done to help the user defining the next date for new measurements.

4 Server Solutions

For the whole system a PHP – framework is provided and a connection from this so defined graphical input/output – interface to a Matlab Webserver is done.

The decision for the use of this two environments is based on the following: on the one hand we have restrictions within the use of two systems instead of one system (e.g. JAVA, …), but on the other side PHP provides all necessary features for an adequate web representation and Matlab [4] is a well validated computer numeric and algebra system. That is why we can more ore less trust the numerical results so far, admittedly that we choose the right model structure and know about the numerical stability of the used algorithms.

The input of the measurements and the optimization of the parameters used in the equation is implemented using the computer numeric/-algebra package Matlab and can be controlled with the included Webserver connection to a graphical user interface. This assures the numerical reliability of the used mathematical method.

The advantage of using the graphical interface and using PHP for solving the defined concept is that it can be easily used for modular extensions. That's the background for further approaches, implementing new and different models and extensions of the described application.

A complex question was to implement the user interface in such a way that physicians can work with the tool both efficiently and in a responsibly way. To reach this goal a comparison with a linear growth was integrated into the tool as a polynomial smoothing function of first order. In principle a difficulty of the project was the precise definition and specification of the model assumptions and the model results. The result is of course defined by the model assumptions. Comparison to the linear result was here a determining part of implementing a practical user's tool [Fig. 3].

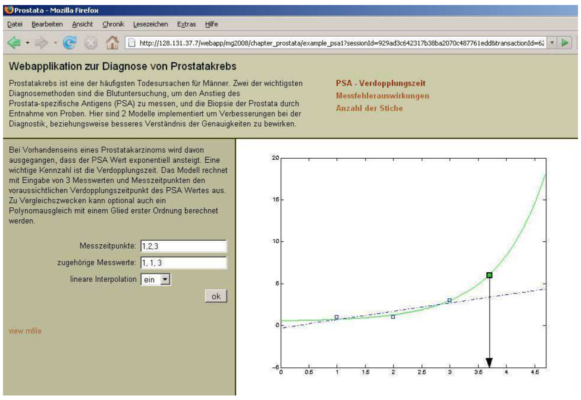

Fig. 3 Exponential fitting model assumption versus Linear Fit

The graphical interface plots the results, shows the doubling time with a black arrow and compares this computation with a linear growth as the picture above shows [Fig. 3]. As we could see in the discussions implementing this tool, a main aspect for physicians is to see whether the results are realistic or not and – most important – that they can compare it with other used "models" for the growth of the PSA marker. (in this case a "normal" linear growth. As mathematical models are still kind of "exotic" in the daily use in medicine comparing the results to studies as [5], which are the main origin of concepts and information is important.

To raise the reliability of the tool in everyday use, a main goal was to implement the error estimation for every patients use. Thus the doctor can visualize immediately, not only the result itself, but also the reliability of the - for him possibly astonishing - result.

So directly within the application it can be shown that by rising the number of measurements also the relevance of the results increases. In practical application this is very pleasant for the treating doctor, because mostly dealing with patients from 50 regularly precaution investigations should be done, and so after a few appointments a relatively high prediction exactness appears. Always assumed that the model of exponential development is accepted - as described above – i.e. that the appearance of a possible cancer illness is already implemented in the selection of the used function.

5 Other Applications & Summary

A medical advantage of the PSA analysis comparing with other methods in prostate cancer detection is the easy way to receive data by blood tests. The second model - which is developed at the moment – has another goal. That is to accomplish an invasive intervention as carefully and as short as it can reasonable done [6].

The prostate gland can measure between 20 cm³ (average value) and 100 cm³. To see wether a patient has a prostate cancer or not, samples are taken via a needle out of the prostate gland. The cancer is assumed to be evenly distributed in the tissue but can be segmented in different parts. The stitches are done evenly distributed over the whole volume of the prostate gland. The only possibility to reach the prostate gland is via the colon [Fig. 4]. The stitches are done following a line, doing always two stitches with an angle smaller than 90° [Fig. 5]. This check up is done in Austria about 100 times every day.

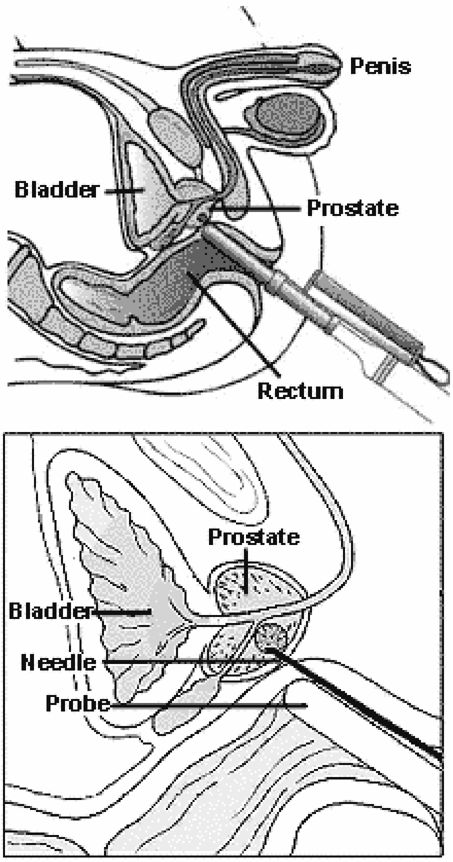

Fig. 4 Analysis of the Prostate Gland

The question is how many stitches are needed with a given volume of the prostate gland to reach a reliability of the result higher than a given percentage (of about 90%-95%). This computation is very useful as the goal is to minimize the number of stitches. This leads to a significant pain reduction for the patient.

This second structure implemented in the PHP framework is a simplified 3D representation of a prostate. For calculations it is assumed that the prostate has the shape of a rotational ellipsoid. The size of this profile is detected by supersonic measures and a tumour is supposed to be a sphere with a definable diameter.

For physicians it is of main interest how many punching tests are important to get a result with a well defined probability to detect a tumour in the gland.

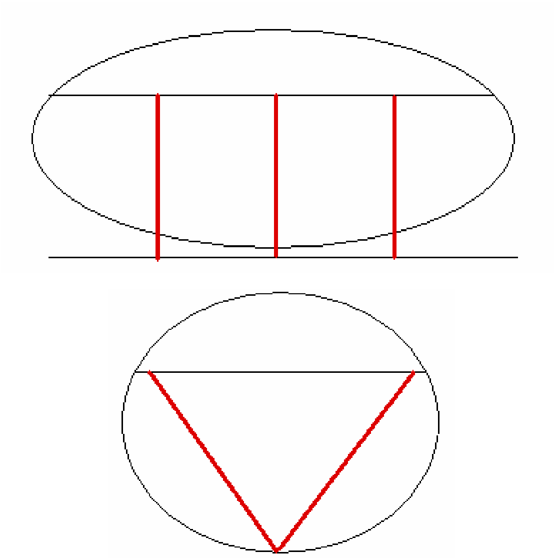

Fig. 5 Predefined Direction of the stitches shown in longitudinal section and in cross section

All calculations are done with numerical randomly taking out test cylinders of the whole volume; as the results can not be calculated with Buffon's Nail Principle [7] we have to compare the tested volumes with both, the total volume of the prostate gland and the reachable volume. The values calculated are the testable area (note that not the whole prostate can be tested because the probe needle is not long enough), and the probabilities to detect a tumor in the tested area and in the whole area.

As medical studies show it is important not to deal with too small volumes of tumor tissue. As – for older patients – most of the tested persons can have tissue in the prostate gland that can be classified as cancer tissue dealing with small volumes could show false results.

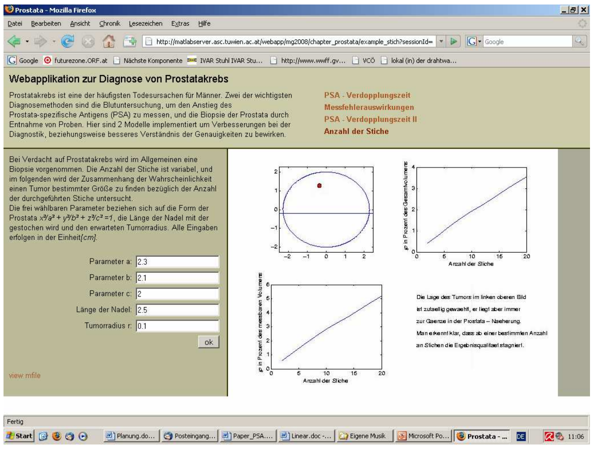

Fig. 6 Linear behavior for small diameters of a carcinoma

As we assume the cancer tissue to have a given volume an interesting behavior appears. As the results for small areas to be found in the prostate gland is nearly linear as we can see [Fig.6] – so that every stitch would improve the reliability of the result - the behavior for bigger volumes to be found becomes a saturated curve. [Fig. 7]

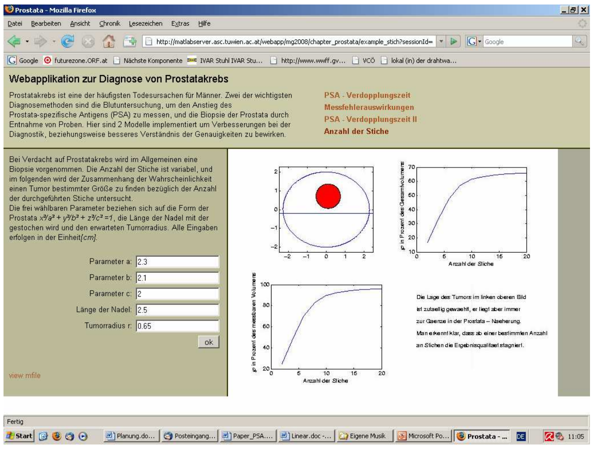

Fig. 7 Saturated behavior

We see that after a well defined number of tests the solution does not become significantly better. This is an important experience, because this sort of examination is quite painful for the patient.

So, with this model for every days use and for educational reasons two things are important: On one hand the size of the prostate gland in relation to the length of the needle. By implementing the model it can be made clear to the physicians that in most cases, because the length of the needle is shorter than the diameter of the prostate gland – it is not possible to reach all areas of the prostate gland.

Secondly, it can be made clear, that the solution changes with increasing size of the assumed tumor from a linear behaviour to a quickly saturated one. So as we assume that we are looking for a not marginal tumor, the likelihood of finding the tumor rises very fast on more than 90 percent of the volume that can be reached and measured. This knowledge can save the patient up to 10 stitches. Indeed – as described above - it becomes also clear that one can reach - in this example - only about 70% at all, because not all areas are accessible determined by the choice of the method.

If the first application introduced teaching areas are applicable as well as diagnostics. The second tool rather is a teaching application, which could help to bring understanding of fundamental relations and parameter values to physicians. With both methods and implementations the challenge was to create a substantial educational and diagnostic model, the realization as a web tool and the narrow collaboration between physiologists and simulation experts to make the tools as applicable as possible.

6 References

[1] R.G. Nixon, M.H. Wener, K.M. Smith, R.E. Parson, S.A. Strobel, M.K. Brawer. Biological Variation of Prostate Specific Antigen Levels in Serum: an Evaluation of day-to-day physiological Fluctuations in a well-defined cohort of 24 Patients. *The Journal of Urology*, Vol.157, 2183-2190, 1997.

[2] http://www.pcngcincinnati.org/psa/calculator.html

[3] R. Choo, L. Klotz, G. Deboer, C. Danjoux, G.C. Morton. Wide Variation of Prostate-Specific Antigen Doubling Time of untreated, clinically localized, low-to-intermediate grade, prostate carcinoma, *BJU International*, 94, 295-298, 2004.

[4] Mathworks. *Using Matlab version 5*. The Mathworks Inc., Natick, 1998.

[5] G.J. Kelloff, D.S. Coffey, B.A.Chabner, A.P.Dicker, K.Z.Guyton, P.D.Nisen, H.R.Soule, A.V. D'Amico, Prostate-Specific Antigen Doubling Time as a Surrogate Marker for Evaluation of Oncologic Drugs to Treat Prostate Cancer. *Clinical Cancer Research*. Vol. 10, 3927-3933, 2004

[6] A.R. Vashi, K.J. Wojno, B. Gillespie, J.E. Oesterling. A Model for the Number of Cores per Prostate Biopsy based on Patient Age and Prostate Gland Volume. *The Journal of Urology,* Vol.159, 920-924, 1998

[7] D.A. Klain, ,G. Rota. Introduction To Geometric Probability, Cambridge University Press, 2006

