

УДК 681.2

Дослідження регресійних моделей динаміки глікемії в умовах вимірювальної невизначеності

Чмихова^f О. В., ORCID [0000-0002-9198-9701](https://orcid.org/0000-0002-9198-9701)Шапов^s П. Ф., д.т.н. проф., ORCID [0000-0003-1917-0790](https://orcid.org/0000-0003-1917-0790)

Кафедри промислової і біомедичної електроніки

Національний технічний університет «Харківський політехнічний інститут» kpi.kharkov.ua

Харків, Україна

Павлюк В. М., лікар вищої категорії, ORCID [0000-0003-0178-0814](https://orcid.org/0000-0003-0178-0814)

Відділ урології

БКДЦ Milner-Medical, [milner-medical](http://milner-medical.com)

Харків, Україна

Севрюкова О. С., інженер, ORCID [0000-0003-0047-2027](https://orcid.org/0000-0003-0047-2027)

Навчальний відділ

Національний технічний університет «Харківський політехнічний інститут» kpi.kharkov.ua

Харків, Україна

Анотація—У статті розглядається статистичне обґрунтування процедур ідентифікації параметрів моделі динаміки глікемії в умовах невизначеності результатів вимірювань. Обґрунтовано вибір регресійної моделі з випадковими коефіцієнтами, що характеризує зміни рівня глікемії при істотних обмеженнях на обсяг первинної вимірювальної інформації. Проаналізовані існуючі методи визначення діабету 2 типу, а також моделі для ідентифікації діабетичних станів. Визначені три моделі для яких можна оцінити похибки вимірювання елементів множин. Їм слід надати перевагу при ідентифікації діабетичного стану. Також виявлено одну основну модель, яка найбільш прийнятна для задач розпізнавання параметричної моделі динаміки глікемії при перевірці справедливості статистичних гіпотез.

Ключові слова — динаміка глікемії; регресійна модель; ефективність ідентифікації; поліноміальна модель; параметрична невизначеність.

I. ВСТУП

Наразі існує багато методів інвазійного визначення глікемії, які мають широке розповсюдження. Однак, такі методи можуть привести до інфекційних захворювань. Тому зараз набирають популярність дослідження неінвазійного метода визначення глікемії, які можна буде використовувати у медичних цілях. Вже досліджено кілька можливих методів, які зможуть точно визначити рівень глюкози в крові без пошкодження шкіряного покриву:

- 1) Оптичний метод заснований на зміні кольору шкіри. Для цього використовуються пучки різної довжини. Прилад сканує відбиток чи поглинання пучків, після чого видає результат. Однак такий метод вимагає ретельного калібрування, інакше точність даних може стати сумнівною [1].

- 2) Також на піку досліджень знаходиться лазерна спектроскопія. В основі метода лежить використання інфрачервоного джерела світла і дифузійного відбивача. Контакт шкіри відбувається з рівною стороною напівсферичної лінзи, що добре впливає на записи спектра шкіри. При цьому з верхніх шарів шкіри збираються розсіяні фотони, фокусуються на детекторі, після чого відбувається аналіз зібраних даних. Завдяки тому, що верхні шари шкіри можуть накопичувати глюкозу в капілярах, за коефіцієнтом поглинання певних довжин хвиль можна визначити її кількість за допомогою спеціального багатовимірного аналізу даних [2].
- 3) Все більше набирає популярність метод визначення перших симптомів діабету по варіабельності серцевого ритму. Вже було



проведено кілька досліджень, завдяки яким було встановлено, що є залежність між діабетом і частотою серцевих скорочень. Для цього пацієнтам з діабетом було запропоновано носити пристрої для моніторингу роботи серця. Після чого порівнювалися дані моніторингу при гіпоглікемічних подіях [3].

II. ПОСТАНОВКА ПРОБЛЕМИ

Статтю присвячено проблемі підвищення ефективності ідентифікації параметрів моделі динаміки глікемії за мінімально обмеженим обсягом результатів глюкозотолерантного тестування. Неоднорідність динаміки глікемії забезпечується, при цьому, фактором стимулювання секреції інсуліну екзогенною глюкозою (глюкозним навантаженням). Для вирішення проблеми аналітично досліджуються параметричні підходи до відновлення поліноміальних регресійних моделей по обмеженим експериментальним даним з узагальненням на функціональні особливості динаміки глікемії. Такі підходи забезпечують первинну вимірвальну біохімічну інформацію щодо змін математичного очікування рівня цукру в крові, як функції часу.

Ця проблема пов'язана, з одного боку, з апріорною нестационарністю таких залежностей (як експериментально індукованих перехідних процесів), а з іншого боку, з непереборними обмеженнями на обсяги результатів вимірювань в неповних (усічених) експериментах.

III. АНАЛІЗ ЛІТЕРАТУРИ

Відомо, що параметричні підходи застосовуваних методів статистичного аналізу [4]–[7] можуть бути неадекватними, викликаючи систематичні локальні і глобальні зміщення модельних функціональних залежностей.

IV. МЕТА

Обґрунтування вибору регресійної моделі зміни рівня глікемії при істотних обмеженнях на обсяг первинної вимірвальної інформації, які обумовлюють невизначеність нелінійних змін моделі при нормативних глюкозних навантаженнях.

V. АНАЛІЗ ВАРІАНТІВ ВІДНОВЛЕННЯ ПОЛІНОМІАЛЬНИХ РЕГРЕСІЙ В НЕПОВНИХ ВИМІРОВАЛЬНИХ ЕКСПЕРИМЕНТАХ

Введемо дві множини вимірюваних величин:

X – моменти часу, що визначають послідовності вимірювань рівня глюкози;

Y – результати вимірювання рівнів глюкози у фіксовані моменти часу.

Відомо [8], що дві множини елементів X і Y пов'язані функціональною залежністю, якщо кожному елементу $x \in X$ може бути поставлений в однозначну відповідність елемент $y \in Y$. Ця залежність називається функцією, якщо множина X – вектори, а множина Y – скаляри. Однак, існують і такі

залежності, де кожному вектору X ставиться у відповідність число y , отримане за допомогою випадкового випробування, згідно умовної щільності $P(y|x)$. Інакше кажучи, кожному x ставиться у відповідність закон $P(y|x)$, згідно з яким у випадковому випробуванні реалізується вибір y .

Існування такого роду залежностей відображає наявність стохастичних зв'язків між вектором X і скаляром y . Повне знання стохастичних зв'язків вимагає відновлення умовної щільності $P(y|x)$. Завдання ж відновлення умовної щільності надзвичайно важке. Однак, часто на практиці (наприклад, в задачах обробки результатів вимірювання) потрібно знати не щільність, $P(y|x)$ а лише одну з її характеристик: функцію умовного математичного очікування, тобто функцію, яка кожному x ставить у відповідність число $y(x)$, що дорівнює математичному очікуванню скаляра y

$$y(x) = \int yP(y|x)dy$$

Функція $y(x)$ є регресією, а завдання її відновлення у вигляді оцінки функції умовного математичного очікування — завданням відновлення регресії.

Задача відновлення регресії є однією з основних задач прикладної статистики. До неї зводиться проблема інтерпретації результатів прямих експериментів. Нехай закономірність впливу часу t на функцію відгуку $Y(t)$ пов'язує математичне очікування Y з вимірюваними значеннями величини часу t .

Будемо розглядати Y як контрольований параметр, залежність якого від t – невідома і є стохастичною.

Оцінка контрольованого параметра Y проводиться на підставі результатів вимірювання значень величин t з подальшим функціональним перетворенням цієї величини в оцінку Y^* .

Розглянемо в загальному вигляді залежність параметра Y від величини t для набору t_1, \dots, t_m моментів часу t

$$Y_i = F(t_i), \quad i = \overline{1, m} \quad (1)$$

При відсутності інформації про вид фізичної моделі для вибору функціонального зв'язку $F(\cdot)$, такий зв'язок передбачається стохастичним [9], [10], для якого залежність (1) для моменту часу t_i формально може бути записана у вигляді множинної лінійної регресії \hat{Y}_i порядку L регресорами, в загальному випадку залежними від часу

$$Y_i = \hat{Y}_i + \varepsilon_i,$$



де

$$\hat{Y}_i = \sum_{r=0}^L b_r t_i^r, \quad (2)$$

де b_0, \dots, b_L – коефіцієнти (параметри) регресійної моделі; t_i – регресори $i = \overline{1, m}$; ε_i – випадковий залишок результату обчислення значення Y_i , що задовольняє умовам взаємозалежності та нормальності розподілу імовірностей [4], [11]; L – порядок моделі ($L = m - 1$).

Дослідження впливу можливих варіантів невизначеності в параметрах моделі (2) на точність обчислення функцій \hat{Y}_i і Y_i показує, що таку модель найзручніше розглядати як випадкову (з випадковими регресорами).

Існує кілька варіантів моделі (2) з випадковими регресорами (Таблиця 1).

VI. АНАЛІЗ НЕВИЗНАЧЕНОСТІ ДОСЛІДЖУВАНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ

У функціональній моделі випадковим є регресор t_i (через похибки його вимірювання δ_i).

Моделі з помилками у змінних — це варіант функціональної моделі, в якій випадковий фактор пов'язаний не тільки з похибками δ_i , але і з помилками визначення значень функції Y .

У структурній моделі змінні Y і (t_i) – випадкові, а зв'язок, у вигляді апіорно невідомої функції $F(\cdot)$ існує не між математичними очікуваннями цих змінних, а між їх випадковими значеннями з невизначеністю для порядку L їх регресійного зв'язку.

Моделі регресії другого роду, яка називається ще моделлю компонент дисперсій, є різновидом структурної моделі, в якій крім Y і (t_i) випадковими є і коефіцієнти β_0, \dots, β_L .

Моделі контрольованих змінних — це варіант попередньої моделі, проте тут, хоча регресори і випадкові, середні значення цих регресорів контролюються.

Можна показати [8], [12], що задача відновлення регресії фактично зводиться до схеми мінімізації середнього ризику, в якій, на відміну від задачі навчання розпізнавання образів, функція (2) може приймати будь-які значення з інтервалу $(-\infty; \infty)$. Наявність мінімуму ризику вказує на досяжність, при обмеженнях на невизначеність, що виникає випадкового залишку ε і похибок δ_i , Δ_i , якщо порядок регресії L – фіксований, а сама модель Y динаміки глікемії відноситься до варіантів функціональної моделі або моделі з помилками в змінних.

Таблиця 1 Поліноміальні моделі функції $Y(t)$ за видами параметричної невизначеності

№	Найменування моделі	Математична реалізація функції (1)	Невизначеності параметрів регресії (2)	Контрольовані елементи моделі
1	Функціональна	$M(Y_i) = \sum_{r=0}^L \beta_r (t_i + \delta_i)^r$	δ_i – похибка вимірювання t_i ($i = \overline{1, m}$)	$M(Y_i), \beta_r, L$
2	З помилками у змінних	$M(Y_i + \varepsilon_i) = \sum_{r=0}^L \beta_r (t_i + \delta_i)^r$	δ_i, ε_i – похибка обчислення функції Y_i	β_r, L
3	Структурна (умовна регресія)	$M \left(Y_i + \varepsilon_i \mid \left[\{t_i + \delta_i\}^r \right]_{r=0}^{L^*} \right) = \sum_{r=0}^L \beta_r (t_i + \delta_i)^r$	$\delta_i, \varepsilon_i, L^*$ – невизначеність у виборі порядку L моделі (2)	$M(\beta_r)$
4	Компонент дисперсій	$M \left(Y_i + \varepsilon_i \mid \left[\{t_i + \delta_i\}^r \right]_{r=0}^{L^*} \right) = \sum_{r=0}^L (\beta_r + \Delta_r) (t_i + \delta_i)^r$	$\delta_i, \varepsilon_i, L^*, \Delta_r$ – похибка визначення коефіцієнтів регресії (2)	Дисперсії $\delta_{t_i}^2$ і $\delta_{Y_i}^2$
5	Контрольованих змінних	$M \left(Y_i + \varepsilon_i \mid \left[\{t_i + \delta_i\}^r \right]_{r=0}^{L^*} \right) = \sum_{r=0}^L (\beta_r + \Delta_r) M \left([t_i + \delta_i]^r \right)$	$\delta_i; \varepsilon_i; L^*; \Delta_r$	$M(t_i)$



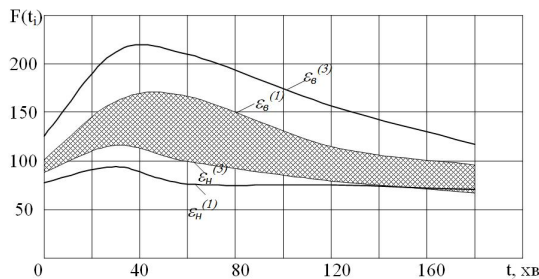


Рис. 1 Довірчі інтервали для біохімічних параметрів за станами Θ_1 і Θ_2 (довірча імовірність – 0,95)

Таблиця 2 Оптимальні значення \bar{R}_p^2 для L та виду регресійних моделей ($N = 30$; $m = 6$)

$\bar{R}_{p \max}^2$	L	R_p^2	НОМЕР РЕГРЕСІЙНИХ МОДЕЛЕЙ
0,724	5	0,711	1
0,801	4	0,793	2
0,656	3	0,644	3

На рис. 1 представлені, як приклад оцінювання поліноміальних регресій, що моделюють невизначеність зміни рівнів глікемії для альтернативних станів (модель № 2 Таблиця 1):

- Θ_1 – норма (відсутність діабету);
- Θ_2 – діабет 2-го типу.

Для цих станів були розраховані нижня ε_n і верхня ε_e межа контрольних інтервалів (при довірчій імовірності $P = 0,95$), що ілюструють появу невизначеності (затонована область) при ідентифікації даних станів.

Ці межі обчислювалися відповідно до моделі інтервального оцінювання параметрів при відомій дисперсії їх точкових оцінок [10]

$$\varepsilon_n = \bar{\varepsilon}(T) - 2\sigma_T,$$

$$\varepsilon_e = \bar{\varepsilon}(T) + 2\sigma_T,$$

де $\bar{\varepsilon}(T)$ – поліноміальна модель 4-го порядку для математичного очікування процесу зміни глюкози для часу спостереження T (для обраного біологічного стану Θ_i , де t приймає одне із значень – 1 або 3); σ_T – стандартне відхилення для результату вимірів рівня глюкози.

З рис. 1 видно, що контрольні інтервали для 1 і 2 завжди перетинаються, що обумовлює високу невизначеність прийнятих діагностичних рішень при класифікації моделей динаміки глікемії (для рівня значущості $\alpha = 0,05$).

Для вибору регресійної моделі, що має мінімальне перетинання між довірчими інтервалами рис. 1, можна використати два підходи [9]:

- обчислення оцінок середнього ризику для вибраних станів діабетичних порушень для кожної із регресійних моделей (вибір по мінімуму ризику);
- обчислення скорегованого коефіцієнту детермінації, що характеризує точність регресійної моделі (вибір по максимуму скорегованого коефіцієнту).

В табл. 2 надані результати обчислень максимальних значень скорегованого коефіцієнту детермінації для перших трьох моделей з табл. 1.

Для розрахунку скорегованого коефіцієнту детермінації використовувалося рівняння.

Скоригований коефіцієнт детермінації

$$\bar{R}_p^2 = 1 - \left(1 - R_p^2\right) \left(\frac{n}{n-p}\right)$$

R_p^2 — виборочний множественный коэффициент корреляции; $p = L + 1$, $n = N * m$, N_θ кількість пацієнтів, що знаходяться у стані θ .

Результати обчислень, що надані в Таблиця 2, вказують на те, що модель 2 має більшу точність, оскільки її скорегований коефіцієнт детермінації є найбільшим з усіх можливих.

ВИСНОВКИ

- 1) Процедура відновлення поліноміальної регресії, як моделі динаміки глікемії, спрощується, якщо кількість можливих варіантів цієї моделі – кінцева. У цьому випадку відновлення еквівалентно ідентифікації параметрів поліноміальної регресії для кінцевої множини параметрів, отриманих на етапі навчання системи ідентифікації.
- 2) При використанні моделей Таблиця 1 для ідентифікації діабетичних станів перевагу слід віддати першим трьом моделям, для яких можна оцінити похибки вимірювання елементів множин як X так і Y . Моделі № 4 і № 5 мають, на жаль, непереборну невизначеність в кутових коефіцієнтах регресії, що призводить до неефективності оцінювання математичного очікування елементів множини Y .
- 3) Використання моделі № 2 найбільш прийнятне для задач розпізнавання параметричної моделі динаміки глікемії при перевірці справедливості статистичних гіпотез, кількість яких відповідає числу діагностованих діабетичних станів.

ВНЕСОК АВТОРІВ

Чмихова О. В. — виконавець роботи, Щапов П. Ф. — науковий керівник подання, Павлюк В. М. — кваліфікований консультант, Севрюкова О. С. — аналітична робота.

ПЕРЕДІК ПОСИЛАНЬ

- [1] “Combo Glucometer (CoG)- Non-invasive Hybrid Glucometer.” [Online]. Available: <https://cnogacare.co/products/cog-info/>.
- [2] K. Laska, “Blood Sugar Monitoring: Using Infrared Instead of Invasive Techniques,” *MEDICA-tradefair.com*, URL: [https://www.medicatradefair.com/en/News/Interviews/Previous Interviews/Interview s 2019/Blood Sugar Monitoring Using Infrared Instead of Invasive Techniques](https://www.medicatradefair.com/en/News/Interviews/Previous%20Interviews/Interview%20s%2019/Blood%20Sugar%20Monitoring%20Using%20Infrared%20Instead%20of%20Invasive%20Techniques).
- [3] D. Muoio, “Study finds wearable heart rate monitors feasible for early detection of hypoglycemia,” *Mobi Health News*. [Online]. Available: <https://www.mobihealthnews.com/content/study-finds-wearable-heart-rate-monitors-feasible-early-detection-hypoglycemia>.
- [4] N. Dzhonson, *Statistika i planirovaniye eksperimenta v tekhnike i nauke: Metody planirovaniya eksperimenta [Statistics and experimental design in engineering and science: Experimental design methods]*. Moscow: Mir, 1981.
- [5] A. Tsanas, M. A. Little, and P. E. McSharry, “A Methodology for the Analysis of Medical Data,” in *Handbook of Systems and Complexity in Health*, New York, NY: Springer New York, 2013, pp. 113–125, URL: http://link.springer.com/10.1007/978-1-4614-4998-0_7. DOI: 10.1007/978-1-4614-4998-0_7
- [6] V. I. Yunkerov and S. G. Grigor’ev, *Matematiko statisticheskaya obrabotka dannykh meditsinskikh issledovaniy [Mathematical and statistical processing of medical research data]*. St. Petersburg: VMedA, 2002.
- [7] W. G. HOPKINS, S. W. MARSHALL, A. M. BATTERHAM, and J. HANIN, “Progressive Statistics for Studies in Sports Medicine and Exercise Science,” *Med. Sci. Sport. Exerc.*, vol. 41, no. 1, pp. 3–13, Jan. 2009, PMID: 19092709. DOI: 10.1249/MSS.0b013e31818cb278.
- [8] V. N. Vapnik, *Vosstanovleniye zavisimostey po empiricheskim dannym [Dependency Recovery from Empirical Data]*. Moscow: Glavnaya redaktsiya fiziko-matematicheskoy literatury izdatel’stva “Nauka,” 1979.
- [9] D. Serbii, *Lineynyy regressionnyy analiz [Linear Regression Analysis]*. Moscow: Mir, 1980.
- [10] N. V. Smirnov, *Kurs teorii ymorvinostey i matematicheskoy statistiki dlya tekhnicheskikh prilozheniy [Course on the theory of morphine and mathematical statistics for technical applications]*. Moscow: Nauka. Glavnaya redaktsiya Fizmatliteratury, 1969.
- [11] J. S. Armstrong, “Illusions in regression analysis,” *Int. J. Forecast.*, vol. 28, no. 3, pp. 689–694, Jul. 2012, DOI: 10.1016/j.ijforecast.2012.02.001.
- [12] P. F. Shchapov, *Pidvyshchennya dostovirnosti kontrolyu i diahnozyky ob’ektiv v umovakh nevyznachennosti [Improving the reliability of control and diagnostics of objects under uncertainty]*. Kharkiv: KHNADU, 2011.

Надійшла до редакції 01 жовтня 2019 р.



Исследование регрессионных моделей динамики гликемии в условиях измерительной неопределенности

Чмихова^f О. В., ORCID [0000-0002-9198-9701](https://orcid.org/0000-0002-9198-9701)

Шапов^s П. Ф., д.т.н. проф., ORCID [0000-0003-1917-0790](https://orcid.org/0000-0003-1917-0790)

Кафедры промышленной и биомедицинской электроники

Национальный технический университет «Харьковский политехнический институт» kpi.kharkov.ua

Харьков, Украина

Павлюк В. М., врач высшей категории, ORCID [0000-0003-0178-0814](https://orcid.org/0000-0003-0178-0814)

Отдел урологии

БКДЦ Milner-Medical, milner-medical

Харьков, Украина

Севрюкова О. С., инженер, ORCID [0000-0003-0047-2027](https://orcid.org/0000-0003-0047-2027)

Навчальний відділ

Национальный технический университет «Харьковский политехнический институт» kpi.kharkov.ua

Харьков, Украина

Аннотация—В статье рассматривается статистическое обоснование процедур идентификации параметров модели динамики гликемии в условиях неопределенности результатов измерений. Обоснован выбор регрессионной модели со случайными коэффициентами, характеризующий изменения уровня гликемии при существенных ограничениях на объем первичной измерительной информации. Проанализированы существующие методы определения диабета 2 типа, а также модели для идентификации диабета политических состояний. Определены три модели, для которых можно оценить погрешности измерения элементов множеств. Им следует отдать предпочтение при идентификации диабетического состояния. Также выявлено одну основную модель, которая наиболее приемлема для задач распознавания параметрической модели динамики гликемии при проверке справедливости статистически гипотез.

Ключевые слова — динамика гликемии; регрессионная модель; эффективность идентификации; полиномиальная модель; параметрическая неопределенность.



UDC 681.2

Investigation of Regression Models of Glycemic Dynamics under Conditions of Measurement Uncertainty

O. V. Chmykhova^f, ORCID [0000-0002-9198-9701](https://orcid.org/0000-0002-9198-9701)

P. F. Shchapov^g, Dr.Sc.(Eng.) Prof., ORCID [0000-0003-1917-0790](https://orcid.org/0000-0003-1917-0790)

Department of Industrial and Biomedical Electronics

National technical university «Kharkiv Polytechnic institute» kpi.kharkov.ua

Kharkiv, Ukraine

V. M. Pavliuk, Doctor of the highest category, ORCID [0000-0003-0178-0814](https://orcid.org/0000-0003-0178-0814) Department of urology

Milner-Medical, [milner-medical](http://milner-medical.com)

Kharkiv, Ukraine

O. S. Sevriukova, engineer, ORCID [0000-0003-0047-2027](https://orcid.org/0000-0003-0047-2027)

Training department

National technical university «Kharkiv Polytechnic institute» kpi.kharkov.ua

Kharkiv, Ukraine

Abstract—In the course of the research, it was analyzed that there are currently many widespread methods of invasive detection of glycemia. However, such methods can lead to infectious diseases. To minimize the risk of infection, research into the non-invasive method of determining glycemia, which can be used for medical purposes, is gaining popularity. The article describes the most common of these methods currently used.

The article deals with the statistical substantiation of procedures for identifying the parameters of the glycemic dynamics model in the uncertainty of the measurement results. The heterogeneity of glycemic dynamics is ensured by the factor stimulating insulin secretion by exogenous glucose (glucose loading).

The purpose of the article is to substantiate the choice of regression model of change in the level of glycemia with significant restrictions on the volume of primary measurement information, which cause uncertainty of nonlinear changes of the model at normative glucose loads.

For the first time, polynomial regression analysis was applied to identify variants of polynomial regression recovery in incomplete measurement experiments for existing methods of type 2 diabetes, as well as models for the identification of diabetic conditions. Implementation of such dependencies can confirm the presence of stochastic relationships between the factors of influence and response of the organism. Three models are identified that should be prioritized in identifying the diabetic condition. It is shown in the paper that the research problems are reduced to the procedure of polynomial regression restoration as a model of glycemic dynamics depending on the number of possible variants. The analysis of the model shows that the main elements that influence the result are the sum of the regressors with a certain coefficient and the sum of the random remainder of the result of the calculation, which satisfies the conditions of interdependence and normality of probability distribution. One basic model, which is most suitable for the problems of recognizing the parametric model of glycemic dynamics in verifying the validity of statistical hypotheses, was also identified. The paper deals with a polynomial model of function by types of parametric uncertainty regarding the functional relationship between the factors of influence and reaction of the organism. The paper presents, as an example, the estimation of polynomial regressions that model the uncertainty of changes in glycemic levels for alternative conditions.

Keywords — *the dynamics of glycemia; regression model; efficiency of identification; polynomial model; parametric uncertainty;*

