

## 1. ÚVOD

Při řešení mnoha úloh týkajících se analýzy resp. syntézy různých technických systémů se používá metod matematického modelování. Pokud je předem znám vhodný parametrický vztah, který objasňuje chování daného systému, je úlohou odhadnout jeho parametry vhodnými regresními metodami (viz přehled v /1/).

Podle typu modelu mají tyto parametr (nebo alespoň jejich část) fyzikální smysl. V mnoha případech se však stává, že vzhledem ke složitosti daného systému nebo omezeným znalostem o jeho struktuře není možné specifikovat ani třídu parametrických modelů. Pak záleží na tom, proč se vůbec matematický model hledá. Pokud se požaduje pouze znalost "vyhlazených" funkčních hodnot resp. hodnot derivací ve vybraných bodech, je vhodné použití tzv. neparametrických modelů, které vycházejí pouze z požadavku spojitosti ve zvoleném počtu derivací a informací o konkrétních (výběrových) hodnotách vysvětlované (závislé) proměnné  $y$ .

Dnes již existuje celá řada různých postupů pro odhad neparametrické regresní funkce (viz Brunk /2/, Stone /3/, O'Hagan /4/ a Cheng /5/).

V této práci se omezíme na dvě základní skupiny metod. Jsou to :

1. spline vyhlazování
2. modely založené na řešení konvolutorního integrálu

Je ukázáno, že v obou těchto skupinách je nutno volit jistý parametr  $\alpha$ , který je měřítkem mezi hladkostí a stupněm přiblížení modelu k experimentálním hodnotám. Pro optimální výběr tohoto parametru je vhodné použít metody cross - validation /6/.

## 2. ZÁKLADNÍ POJMY

Mějme k dispozici výběrové body  $(x_i, y_i) \quad i=1, \dots, n$ , které odpovídají modelu měření

$$y_i = g(x_i) + \varepsilon_i \quad (1)$$

Bez újmy na obecnosti je možno provést přečíslování jednotlivých bodů tak, aby platilo, že deterministické abscisy  $\{x_i\}$  jsou uspořádány podle velikosti. Tedy

$$x_1 < x_2 < \dots < x_n$$

Dále se provádí lineární transformace  $\{x_i\}$  do intervalu  $[0,1]$ .

O hodnotách  $y_i$  se předpokládá, že jsou to náhodné nezávislé veličiny pocházející ze symetrického rozdělení s konstantním rozptylem  $\sigma^2 < \infty$  a průměrem  $E(y_i) = g(x_i)$ .

Z rov. (1) pak plyne, že  $\varepsilon_i$  jsou náhodné veličiny s nulovým průměrem a rozptylem  $\sigma^2$ .

Symbolem  $g(x)$  je označena neznámá regresní funkce, o níž se pouze předpokládá, že je lineární kombinací měření  $y_i$  a že je hladká i v prvních derivacích, t.j. pochází ze třídy  $C^2$  funkcí.

(Uveďme, že funkce třídy  $C^m$  je ve svém definičním intervalu spojitá v prvních  $m$  derivacích).

## 3. Metoda cross - validation

Obecná definice postupu cross - validation je uvedena Stonem /6/. Ten také diskutuje její statistické vlastnosti.

V kontextu této práce předpokládejme, že regresní funkce  $g(x)$  závisí na jistém parametru  $\alpha$ , který určuje stupeň vyhlazení (jde o kompromis mezi hladkostí regresního modelu a stupněm přiblížení k experimentálním bodům).

Clark /7/ doporučuje volit optimální  $\hat{\alpha}$  minimalizující speciálně definovanou střední kvadratickou chybu predikce  $C(\alpha)$ , pro kterou platí

$$C(\alpha) = \frac{1}{n} \sum_j \frac{1}{n_j} \sum_{(i \in P_j)} w_i [y_i - \hat{g}(x_i / j, \alpha)]^2 \quad (2)$$

Zde  $P_1, \dots, P_k$  jsou dělení množiny indexů  $N = \{1, 2, \dots, n\}$ ,  $n_j$  je počet prvků v  $P_j$ . Označme  $P_j$  takovou podmnožinu množiny  $N$ , pro kterou platí, že  $P_j \cup P_l = N$ . Pak  $\hat{g}(\cdot / j, \alpha)$  je odhad funkce  $g(\cdot)$  určený na základě bodů  $(x_i, y_i)$   $i \in P_j$ . Symbolem  $w_i$  jsou v rov. (1) označeny statistické "váhy" jednotlivých bodů (platí, že  $w_i > 0$  a  $\sum_{i=1}^n w_i = 1$ ).

Z výše uvedeného plyne, že principem metody cross-validation je minimalizace středních kvadratických chyb predikce v bodech  $(x_i, y_i)$   $i \in P_j$  na základě odhadu  $\hat{g}$  určeného z informací v doplňkových bodech  $(x_i, y_i)$   $i \in P_l$  pro všech  $k$  dělení  $P_1, \dots, P_k$ .

Jednotlivé konkrétní varianty této metody se liší podle toho, jakým způsobem se provádí výběr  $P_1, \dots, P_k$ .

Stone [6] volí  $P_j = \{j\}$  a  $n = k$ . Pak je možno vyjádřit  $C(\alpha)$  ve tvaru

$$C(\alpha) = \frac{n^{-1} \sum_{i=1}^n [y_i - \hat{g}(x_i / \alpha)]^2}{[n^{-1} \text{Tr}(E - A(\alpha))]^2} \quad (3)$$

V rov. (3) je  $\hat{g}(\cdot / \alpha)$  odhad  $g(\cdot)$  získaný na základě všech experimentálních bodů (tedy pro  $k=1$ ,  $P_1 = N$  a  $P_j = \{\emptyset\}$ ). Symbolem  $E$  je označena jednotková matice a  $A(\alpha)$  je  $(n \times n)$  matice pro kterou platí, že

$$[\hat{g}(x_1 / \alpha), \dots, \hat{g}(x_i / \alpha), \dots, \hat{g}(x_n / \alpha)]^T = A(\alpha) [y_1, \dots, y_i, \dots, y_n]^T \quad (4)$$

Symbol  $\text{Tr}(\cdot)$  označuje stopu matice. Rov. (3) definuje klasickou střední kvadratickou chybu predikce (viz [1]).

Je zřejmé, že rov. (3) je (pro případ, kdy je možno definovat  $A(\alpha)$ ) značně výhodná, protože není třeba opakovaně počítat odhady  $\hat{g}(\cdot / j, \alpha)$ .

Pokud se pro výpočet  $C(\alpha)$  používá rov. (2), je možno volit např.  $P_j = \{j, j+10, \dots, j+(k-1)10\}$ , kde  $k = n/10$ . Při použití tohoto postupu vzniká nebezpečí, že  $\hat{g}(\cdot)$  bude záviset na periodicitě indexů vybíraných do  $P_j$ . Z toho důvodu je navrženo konstruovat dělení  $P_j$  takto [7]

$$P_j = \{1 + (j+l) \text{mod}(k) + k \cdot l : l = 0, 1, \dots, L\} \cap N$$

kde  $L = (n-1)/k$ .

Pro velký počet experimentálních bodů ( $n$  řádově ve stovkách) doporučuje Clark [8] volit  $k = 40$  a  $n_j = n/10$ . Přitom se jednotlivé  $P_j$  tvoří na základě náhodného výběru z množiny indexů  $N$ . Detailní popis vlastností  $C(\alpha)$  (s ohledem na jeho použití při odhadu neparametrické regresní funkce) je možno nalézt v práci [12].

#### 4. Spline vyhlazování

V kapitole 2. jsme se omezili na nejčastější případ, kdy se uvažuje, že  $g(\cdot)$  je ze třídy  $C^2$ . Aby bylo možno nalézt odhad  $\hat{g}(\cdot)$ , je třeba definovat jisté požadavky. Pro zajištění dostatečné hladkosti je požadováno, aby  $\hat{g}(x)$  minimalizovala integrál

$$R = \int_{x_1}^{x_n} [g''(x)]^2 dx \quad (5)$$

Stupeň přiblížení modelu  $\hat{g}(\cdot)$  k experimentálním bodům je možno (za předpokladu výše uvedených předpokladů o chybách  $E$ ) vyjádřit vztahem

$$S = \sum_{i=1}^n w_i [y_i - \hat{g}(x_i)]^2 \quad (6)$$

Z teorie spline funkcí je známo, že pro  $S = 0$  vychází  $\hat{g}(x)$  jako interpolační kubický spline. Tento spline se skládá z kubických polynomů definovaných lokálně vždy mezi dvojicí sousedních experimentálních bodů. Tedy

$$\hat{g}(x) = a_i + b_i(x-x_i) + c_i(x-x_i)^2 + d_i(x-x_i)^3 \quad (7)$$

Konstanty  $a_i, b_i, c_i, d_i$  jsou jednoznačně určeny z podmínek  $\hat{g}(x_i) = y_i$  a  $\hat{g}'(x_i^+) = \hat{g}'(x_i^-)$ ;  $\hat{g}''(x_i^+) = \hat{g}''(x_i^-)$ . Jde tedy o požadavek spojitosti  $\hat{g}(\cdot)$  v prvních dvou derivacích (viz [9]). Pro určení spline  $\hat{g}(\cdot)$  v intervalech  $x_{i-1} \leq x \leq x_i$  a  $x_i \leq x \leq x_{i+1}$  je třeba definovat vhodné okrajové podmínky (jejich přehled je uveden v [9]). V dalším budeme

volit tzv. přirozené okrajové podmínky, pro které platí

$$\hat{g}'(x_1) = \hat{g}''(x_1) = \hat{g}'(x_n) = \hat{g}''(x_n) = 0$$

Spline, které splňují tyto podmínky se nazývají přirozené.

V případě, že  $S > 0$  (numerické vyhlazování), je možno s využitím podmínek (5) a (6) hledat odhady  $\hat{g}(\cdot)$  dvěma způsoby.

A.  $\hat{g}(\cdot)$  minimalizuje funkcionál  $R + \alpha S$

kde  $\alpha (0 < \alpha < \infty)$  je parametr určující stupeň vyhlazení. Pro  $\alpha = 0$  jde o interpolační kubický spline a pro  $\alpha \rightarrow \infty$  jde o přímkou vyhlazující experimentální body ve smyslu nejmenších čtverců odchylek.

B.  $\hat{g}(\cdot)$  se vybírá tak, aby byl minimalizován integrál  $R$  vzhledem k omezující podmínce  $S \leq S_0$  pro dané  $S_0 (0 < S_0 < \infty)$ . Parametr  $S_0$  představuje přímo reziduální součet čtverců.

Při  $n \geq 3$  š má úlohu ad A. jednoznačné řešení pro všechna  $\alpha$ . Jak ukázal Schvenberg /10/, je toto řešení kubickým spline. Zároveň platí, že pro malá  $S_0$  existuje parametr  $\alpha = \alpha(S_0)$  takový, že řešení problému ad A. je zároveň řešením ad B. V obou případech však místo podmínky

$\hat{g}(x_i) = y_i$  platí, že

$$\hat{g}'''(x_i^+) - \hat{g}'''(x_i^-) = \alpha w_i [y_i - \hat{g}(x_i)] \quad (8)$$

Reinsch /11/ použil při řešení úlohy ad B. metodu Lagrangeových multiplikátorů (s parametrem  $p$ ) a odvodil pro koeficienty kubického spline (vyjádření viz rov. (7)) vztahy

$$\bar{c} = p Q^T \bar{y} [Q^T W Q + p I]^T \quad (9)$$

$$\bar{a} = \bar{y} - p^{-1} W Q \bar{c} \quad (10)$$

$$d_i = (c_{i+1} - c_i) / (3k_i) \quad i = 1, \dots, n-1 \quad (11)$$

$$b_i = (a_{i+1} - a_i) / k_i - c_i k_i - d_i k_i \quad i = 1, \dots, n-1 \quad (12)$$

V těchto rovnicích je použito označení  $k_i = x_{i+1} - x_i$ ;  $\bar{c} = (c_1, \dots, c_{n-1})^T$ ;  $c_n = c_1 = 0$ ;  $\bar{a} = (a_1, \dots, a_n)^T$ ;  $W$  je diagonální matice s váhami  $w_i$  na diagonále. Matice  $T [(n-2) \times (n-2)]$  je pozitivně definitní tridiagonální matice s prvky

$$t_{ii} = 2(k_i + k_{i+1})/3 \quad t_{i,i+1} = t_{i+1,i} = k_{i+1}/3 \quad (13)$$

Matice  $Q [n \times (n-2)]$  je také tridiagonální s prvky

$$q_{i+1,i} = 1/k_{i+1} \quad q_{ii} = -1/k_i - 1/k_{i+1} \quad q_{i,i+1} = 1/k_{i+1} \quad (14)$$

Pro určení Lagrangeova parametru  $p$  je nutno řešit rovnici

$$F(p) = S_0 \quad (15)$$

kde

$$F(p) = \| W Q (Q^T W Q + p I)^{-1} Q^T \bar{y} \|^2 \quad (16)$$

Symbol  $\| \cdot \|$  zde označuje Eukleidovskou normu.

Reinsch /11/ dokázal, že pro  $p \geq 0$  je  $F(p)$  monotónně klesající konvexní funkce, pro kterou platí, že  $F(p) \rightarrow 0$  při  $p \rightarrow \infty$ . Existuje tedy jeden kladný kořen rov. (15), který se dá určit vhodnou jednorozměrnou minimalizační technikou. Wahba /12/ ukazuje, že Reinschův algoritmus se dá použít i pro řešení úlohy ad A., pokud se dosazuje  $\alpha = 1/p \cdot n$ .

V práci /13/ je dokázáno, že spline vyhlazování odpovídá Bayesovskému odhadu  $\hat{g}(\cdot)$  za předpokladu speciální a priori distribuce  $g(\cdot)$ . Tam jsou také diskutovány další vlastnosti této techniky.

Pro nalezení optimálního  $\hat{g}$  se dá s výhodou použít rov. (3), kde

$$E - A(\alpha) = Q [Q^T W Q + p I]^{-1} Q^T \quad (16)$$

V případě ekvidistantního dělení experimentálních bodů je možno vyjádřit rov. (16) ve zjednodušeném tvaru (viz /12/), což značně usnadňuje výpočty.

Utreras /13/ našel aproximativní výraz pro vlastní čísla matice  $A(\alpha)$  platný pro  $n > 10$ . Pro případ ekvidistantního dělení experimentálních bodů ( $h = x_{i+1} - x_i$  pro libovolné  $i$ ,  $1 < i < n-1$ ) a jednotkovou matici vah  $W = E$  je pak možno psát /14/

$$T_n(A(\alpha)) \cong \sum_{i=1}^n \frac{1}{1 + \frac{\lambda_i^2}{\rho}} \quad (17)$$

kde  $\lambda_1 = \lambda_2 = 0$   $\lambda_i = h(\pi/nh)^4 a_{i-2}$   $i = 3, 4, \dots, n$

a  $\rho$  je Lagrangeův multiplikátor. Konstanty  $a_i$  jsou kladné kořeny rovnice

$$\cos(a_i \pi) \cosh(a_i \pi) = 1 \quad (18)$$

Pro první odhad  $a_i^0$ , je vhodné volit  $a_i^0 = i + 0,5$ . Pokud je  $i > 10$ , je již  $a_i^0$  na 16 platných cifer shodné s  $a_i$ .

Program "Reinsch" pro stolní počítač HP 9825 využívá pro odhad koeficientů vyhlazujícího spli-ne Reinschova algoritmu. V první variantě se volí rozptyl měření  $\sigma^2$  a z něj se počítá  $\xi = n\sigma^2$ . Výstupem je graf znázorňující experimentální body a odhad  $\hat{g}(x/\alpha)$ . Není-li uživatel spokojen, může zadat nové  $\sigma^2$ . V případě, že  $\hat{g}(x/\alpha)$  vyhovuje požadavkům uživatele, počítají se pro jednotlivé body hodnoty  $\hat{g}'(x_i/\alpha)$ ,  $\hat{g}''(x_i/\alpha)$  a  $\int_{x_i}^{x_{i+1}} \hat{g}(x/\alpha) dx$ .

Ve druhé variantě programu "Reinsch" se předpokládá ekvidistantní dělení experimentálních bodů a  $W = E$ . Optimální  $\alpha$  se počítá minimalizací rov. (3), kde  $T_n(A(\alpha))$  je vyjádřeno rov. (17). Pro minimalizaci se používá metody kvadratické aproximace.

Třetí varianta programu se liší od druhé pouze tím, že počítá  $C(\alpha)$  podle Clarkova postupu /8/.

Je tedy vhodná pro neekvidistantní dělení experimentálních bodů příp. různé váhy  $w_i$ .

Výstupy druhé a třetí varianty jsou stejné jako u první varianty. Pro ilustraci činnosti programu "Reinsch" je na obr. 1.a, b, znázorněna funkce  $\hat{g}(\cdot)$  vyhlazující data generovaná ze vztahu

$$y_i = 5 + \sin(12i + 45) + 0,25 \sin(56i) + E_i \quad i = 0,01, 0,02, \dots, 1 \quad (19)$$

V rov. (19) je  $E_i$  náhodná veličina s rozdělením  $N(0, \sigma^2)$ .

Na obr. 1a je voleno  $\sigma^2 = 0,1$  a na obr. 1b je  $\sigma^2 = 1$ . Je zřejmé, že i pro data s větším rozptylem jsou vyhlazující spli-ne  $\hat{g}(\cdot)$  vhodné odhady původní funkce  $g(\cdot)$ . Pouze změny funkčních hodnot, které jsou srovnatelné s náhodnými odchylkami  $E_i$  není pochopitelně možné identifikovat.

## 5. Modely založené na řešení konvolutorního integrálu

Clark /8/ definuje  $\hat{g}(\cdot)$  jako řešení konvolutorního integrálu, definovaného vztahem

$$\hat{g}(x) = \int_{-\infty}^{\infty} h(t) \frac{1}{\alpha} W\left(\frac{x-t}{\alpha}\right) dt \quad (20)$$

Zde  $W(\cdot)$  je zvolená váhová funkce,  $\alpha$  je parametr určující stupeň vyhlazení (platí, že pro  $n \rightarrow \infty$  je  $\alpha \rightarrow 0$ ) a  $h(\cdot)$  je lineární interpolační spli-ne definovaný výrazy

$$h(x) = y_1 \quad \text{pro } x \leq x_1$$

$$h(x) = y_i \frac{(x_{i+1} - x)}{(x_{i+1} - x_i)} + y_{i+1} \frac{(x - x_i)}{(x_{i+1} - x_i)} \quad \text{pro } x_i \leq x \leq x_{i+1} \quad (i = 1, \dots, n-1)$$

$$h(x) = y_n \quad \text{pro } x \geq x_n$$

Váhová funkce  $W(\cdot)$  musí splňovat tyto podmínky :

a)  $W(x) \geq 0$  pro všechna  $x$  a  $W(-x) = W(x)$

b)  $\int_{-\infty}^{\infty} W(x) dx = 1$

c)  $\int_{-\infty}^{\infty} x W(x) dx < \infty$

12.8

$$\sigma^2 = 0.1$$
$$S_0 = 1$$

a)



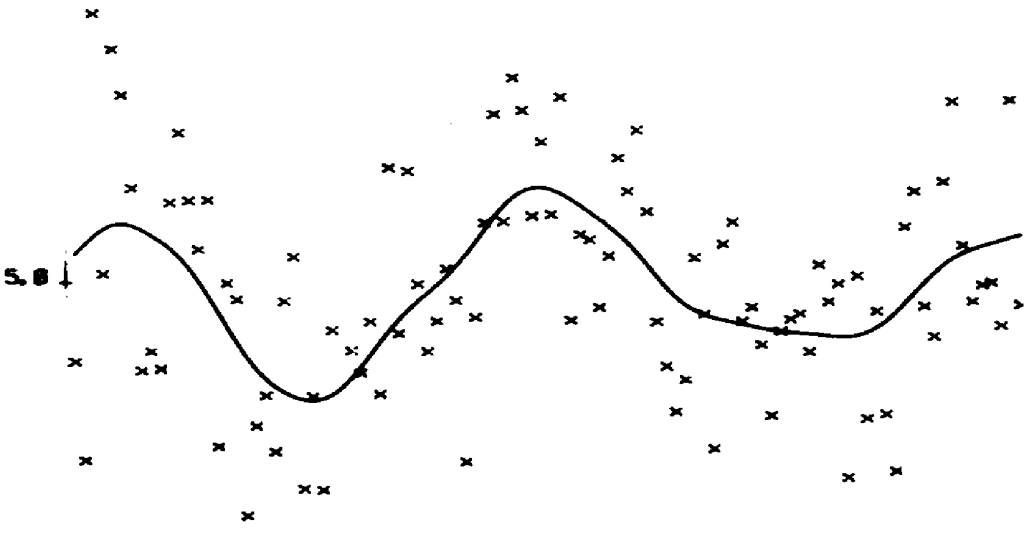
1

2.0  
1.5  
1.0

12.8

$$\sigma^2 = 1$$
$$S_0 = 110$$

b)



2.0  
1.5  
1.0

Obr. čís. 1 : Vyhlasování dat (viz rov. (19)) pomocí programu "Reinsch".

Bendetti /15/ sledovala různé váhové funkce  $W(\cdot)$ . Jako optimální (z hlediska střední kvadratické chyby) vyšla kvadratická funkce typu

$$W(u) = 0,75(1-u^2)_+ \quad (21)$$

kde  $(u)_+ = u$  pro  $u > 0$  a  $(u)_+ = 0$  v ostatních případech.

Ruban /16/ určil, že pro velká  $n$  je možno nalézt optimální  $\alpha$  podle vztahu

$$\hat{\alpha} = \Theta^{-1} n^{-1/5}$$

kde  $\Theta$  je konstanta.

Clark /8/ dokázal, že pokud je  $W(\cdot)$  spojitá, má  $\hat{g}(u)$  z rov. (20) spojitě druhé derivace. Odhad  $\hat{g}(u)$  je také invariantní vůči lineární transformaci měřených  $\{x_i\}$ . Nadto platí, že pokud je  $g(u)$  lineární, představuje  $\hat{g}(u)$  nevychýlený odhad. Další statistické vlastnosti (konzistenci) odhadu  $\hat{g}(u)$  definovaného rov. (20) uvádí Cheng a lin /5/.

Z výpočetního hlediska je výhodné vyjádřit rov. (20) ve tvaru

$$\hat{g}(u) = \sum_{i=1}^n u_i(u) y_i \quad (22)$$

kde

$$u_i(u) = \int_{-\infty}^{\infty} v_i(t) \frac{1}{\alpha} W\left(\frac{u-t}{\alpha}\right) dt \quad (23)$$

Funkce  $v_i(t)$  plyne z ekvivalentního vyjádření lineárního interpolačního spline  $h(u)$  ve tvaru

$$h(u) = \sum_{i=1}^n v_i(u) y_i \quad (24)$$

Po dosazení z rov. (24) do rov. (23) se vzhledem k lokální definovanosti  $v_i(u)$  rov. (23) zjednoduší na tvar

$$u_i(u) = \int_0^{x_1} \frac{1}{\alpha} W\left(\frac{u-t}{\alpha}\right) dt + \int_{x_1}^{x_2} \left(\frac{x_2-t}{x_2-x_1}\right) \frac{1}{\alpha} W\left(\frac{u-t}{\alpha}\right) dt \quad (i=1) \quad (25a)$$

$$u_i(u) = \int_{x_{i-1}}^{x_i} \left(\frac{t-x_{i-1}}{x_i-x_{i-1}}\right) \frac{1}{\alpha} W\left(\frac{u-t}{\alpha}\right) dt + \int_{x_i}^{x_{i+1}} \left(\frac{x_{i+1}-t}{x_{i+1}-x_i}\right) \frac{1}{\alpha} W\left(\frac{u-t}{\alpha}\right) dt \quad (i=2, \dots, n-1) \quad (25b)$$

a

$$u_i(u) = \int_{x_{n-1}}^{x_n} \left(\frac{t-x_{n-1}}{x_n-x_{n-1}}\right) \frac{1}{\alpha} W\left(\frac{u-t}{\alpha}\right) dt \quad (i=n) \quad (25c)$$

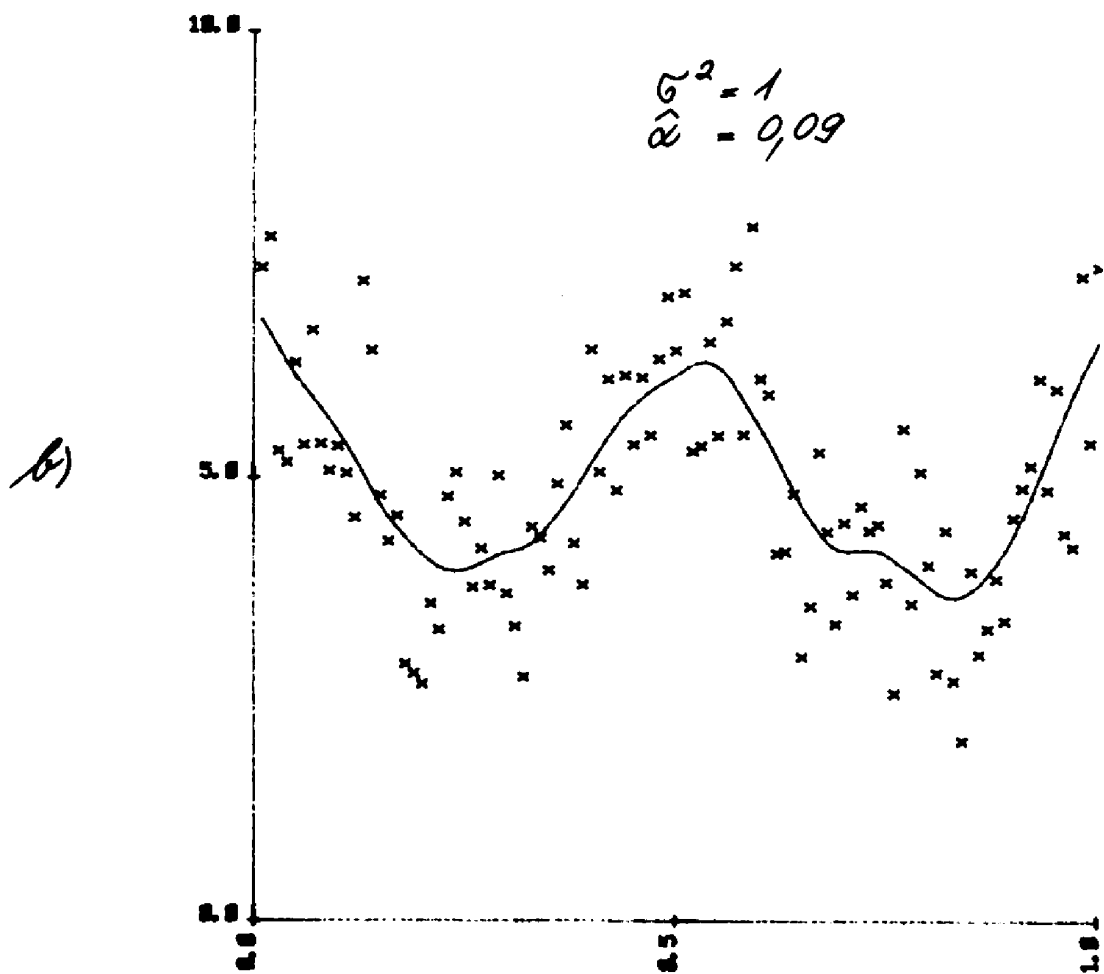
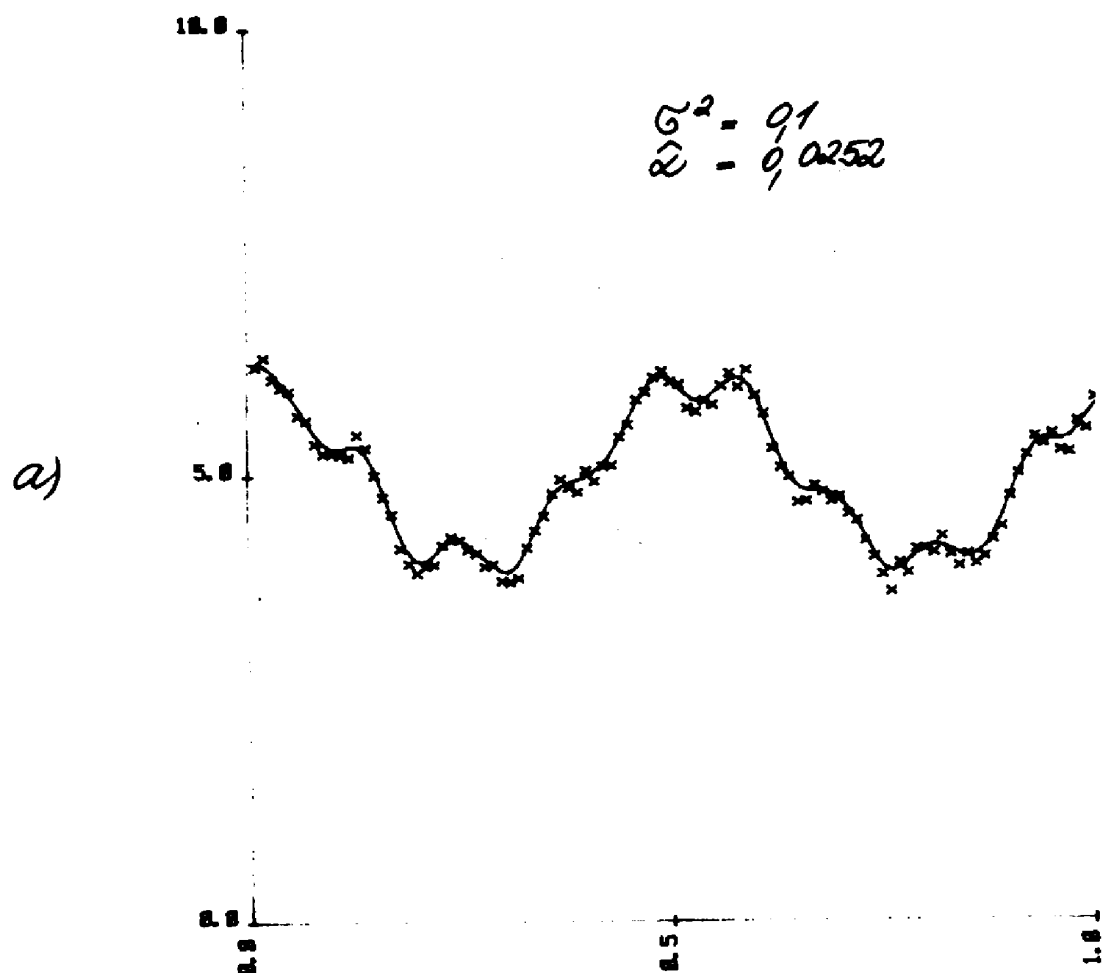
Pokud se volí  $W(\cdot)$  jako kvadratická funkce vyjádřená rov. (21), je poměrně snadné provést analytickou integraci rov. (25a), (25b) a (25c) a nalézt pro  $u_i(u)$  algebraické výrazy (polynomy druhého stupně resp. jejich lineární kombinace).

Pro případ neekvidistantního měření volí Clark parametr  $\alpha = \alpha(u)$  jako funkci počtu opakování a vzdálenosti mezi experimentálními body /7/.

V práci /8/ bylo zjištěno, že odhady  $\hat{g}(u)$  získané z rov. (20) jsou srovnatelné s vyhlazujícími kubickými spline. Jejich výhodou je však to, že nevyžadují řešení soustavy lineárních rovnic ( $u$  které bývá často špatně podmíněná matice koeficientů). Jednotlivá  $u_i(u)$  se počítají přímo a jsou nadto nenulové pouze v jistém omezeném intervalu. Program "Clark" pro stolní počítač HP 9825 vychází při konstrukci  $\hat{g}(\cdot)$  z rov. (22). Je volena kvadratická váhová funkce  $W(\cdot)$  definovaná rov. (21). Výrazy (25a), (25b) a (25c) jsou vyčísleny analyticky, takže se výpočet  $u_i(u)$  omezuje na pouhé dosazování do polynomických výrazů.

Základní varianta tohoto programu vychází ze zadaného parametru  $\alpha$ . Výstupem je opět graf, znázorňující experimentální body a model  $\hat{g}(\cdot)$ . Také ostatní výstupy jsou stejné jako u programu "Reinsch".

Druhá varianta programu umožňuje nalezení optimálního  $\alpha$  minimalizujícího  $C(\alpha)$  definovaného rov. (2). Dělení indexové množiny  $N$  se provádí podle postupu doporučeného Clarkem /8/.



Obr. čis. 2 : Vyhlasování dat (generovaných rov. (19)) pomocí programu "Clark"

Pro minimalizaci  $C(\alpha)$  se opět používá kvadratické aproximace.

Pro ilustraci činnosti programu "Clark" je na obr. 2 a, b znázorněna funkce  $\hat{g}(\cdot)$  vyhlazující data generovaná rov. (19). Na obr. 2 a) je voleno  $\sigma^2 = 0,1$  a na obr. 2 b) je  $\sigma^2 = 1$ . Porovnáním obr. 1 s obr. 2 je zřejmé, že v obou případech jsou odhady  $\hat{g}(\cdot)$  blízké.

## 6. ZÁVĚR

V příspěvku byly diskutovány dvě skupiny metod pro odhad neparametrické regresní funkce. V literatuře se vyskytuje ještě celá řada alternativních postupů (např. neparametrické modely vycházející přímo z obecné definice regresní funkce /17/). Zde uvedené metody však mají velmi dobré statistické vlastnosti, a umožňují nalezení  $\hat{g}(\cdot)$  ze třídy  $C^m$  funkcí. To je výhodné zejména v případech, kdy je účelem stanovit průběh derivace neparametrické regresní funkce. Popsané programy jsou poměrně jednoduché a lze je snadno převést i na jiné počítače. Výhodné je používat varianty umožňující nalezení optimálního parametru  $\alpha$  postupem cross - validation.

## LITERATURA

- /1/ Militký J.: Regresní postupy pro stolní počítač řady HP, skripta pro kurs DT ČsVTS Ostrava, únor 1982
- /2/ Brunk H.D.: Commun. Statist. A 9, 1101 (1980)
- /3/ Stone CH.J. : The Annals of Statist. 5, 595 (1977)
- /4/ O'Hagan A. : J.R.Statist. Soc. B 40, 1 (1978)
- /5/ Cheng K.F., Lin P.E.: Z. Wahrscheinlichkeitstheorie verw. Gebiete 57, 223 (1981)
- /6/ Stone M.: J.R. Statist. Soc. B 36, 111 (1974)
- /7/ Clark R.M.: J.R. Statist. Soc. A 143, 177 (1980)
- /8/ Clark R.M. : J.R. Statist. Soc. B 39, 107 (1977)
- /9/ Ahlberg J.H., Nilson E.N., Walsch J.L. : The theory of splines and their application, Academic Press, New York 1967
- /10/ Schvenberg I.J.: Proc. Math. Acad. Sci. (U.S.A.) 52, 947 (1964)
- /11/ Reinsch CH.H.: Numer. Math. 10, 177 (1967)
- /12/ Craven P., Wahba G.: Numer. Math. 31, 377 (1979)
- /13/ Utreras F.D.: in Smoothing Techniques for Curve Estimation, Lect. Notes in Mathematics No. 757, 1979
- /14/ Utreras F.D.: A Package for Smoothing Noisy Data with Splines, Informe tecnico MA - 80 - B - 209, Universidad de Chile, Mayo 1980
- /15/ Bendetti J.K.: J.R. Statist. Soc. B 39, 248 (1977)
- /16/ Ruban A.I.: Avtomatika i telemekhanika, No. 11, 106 (1979)
- /17/ Watson G.S. : Siskya A 26, 359 (1964)