

UNIVERZITA PARDUBICE

Fakulta chemicko-technologická
Katedra analytické chemie

Licenční studium chemometrie na téma

Statistické zpracování dat

Semestrální práce z 8. soustředění

Předmět: 4.4 Aproximace a vyhlazování křivek

Vedoucí licenčního studia: Prof. RNDr. Milan Meloun, DrSc.

Vypracoval: Ing. Radek Novotný

Licenční studium Statistické zpracování experimentálních dat.

Předmět: 4.4 Aproximace a vyhlazování křivek

Přednášející: Prof. Ing. Jiří Militký, CSc.

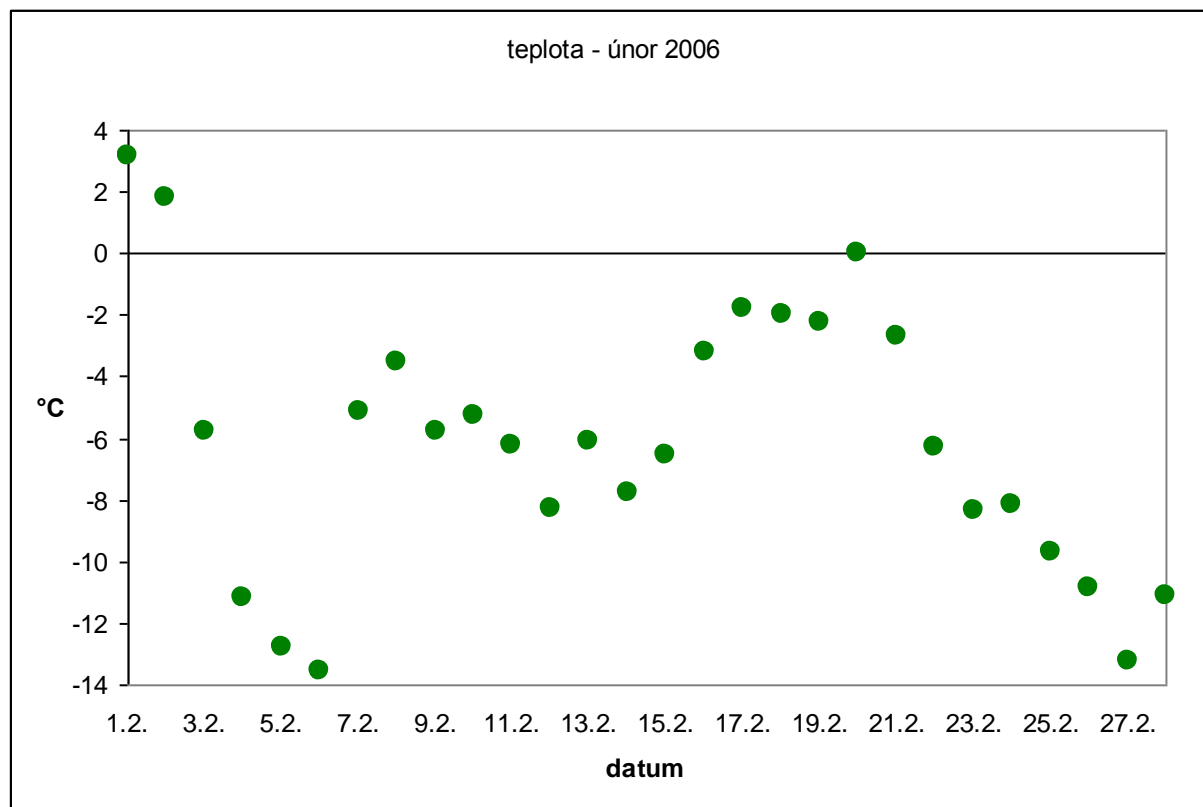
Zadání: Na datech pracoviště nebo datech z internetu zpracovat tři úlohy – aproximace, interpolace, vyhlazování.

Úloha 1: Aproximace experimentálních dat. Jako vstupní data jsou použity průměrné denní teploty v únoru 2006 na výzkumné ploše Q102.

Tabulka 1 – Vstupní data:

datum	1.2	2.2	3.2	4.2	5.2	6.2	7.2	8.2	9.2	10.2	11.2	12.2	13.2	14.2
teplota	3.1	1.8	-5.8	-11.2	-12.8	-13.6	-5.1	-3.6	-5.8	-5.2	-6.3	-8.2	-6.1	-7.7
datum	15.2	16.2	17.2	18.2	19.2	20.2	21.2	22.2	23.2	24.2	25.2	26.2	27.2	28.2
teplota	-6.5	-3.2	-1.8	-2.0	-2.2	0.0	-2.7	-6.3	-8.3	-8.2	-9.7	-10.9	-13.2	-11.1

Obr. 1 – Graf vstupních dat



Úloha byla řešena v programu ADSTAT 1.25. Využita byla metoda lineární, kvadratické a kubické spline regrese se stanovením počtu uzlových bodů od 2 do 5, tedy s konstantními uzlovými intervaly, přičemž byla sledována kvalita proložení naměřených experimentálních hodnot. Dosažené výsledky byly porovnány a bylo zvoleno nejlepší řešení.

Výstupy programu ADSTAT pro úlohu 1:

A) Lineární spline regrese

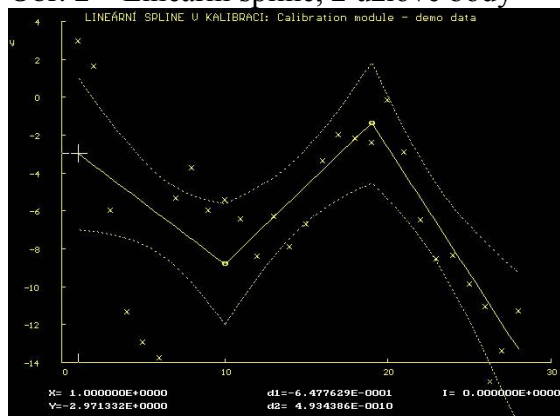
a) Dva uzlové body

Reziduální součet čtverců, RSC:	2.8555E+02
Průměr absolutních hodnot reziduí, Me:	2.3715E+00
Průměr relativních reziduí, Mer[%]:	49.152
Odhad reziduálního rozptylu, $s^2(e)$:	1.1898E+01
Odhad směrodatné odchylky reziduí, $s(e)$:	3.4494E+00

b) Tři uzlové body

Reziduální součet čtverců, RSC:	2.8074E+02
Průměr absolutních hodnot reziduí, Me:	2.5672E+00
Průměr relativních reziduí, Mer[%]:	55.157
Odhad reziduálního rozptylu, $s^2(e)$:	1.2206E+01
Odhad směrodatné odchylky reziduí, $s(e)$:	3.4937E+00

Obr. 2 – Lineární spline, 2 uzlové body



Obr. 3 – Lineární spline, 3 uzlové body



B) Kvadratický spline

a) Dva uzlové body

Reziduální součet čtverců, RSC:	2.2577E+02
Průměr absolutních hodnot reziduí, Me:	2.4057E+00
Průměr relativních reziduí, Mer[%]:	48.819
Odhad reziduálního rozptylu, $s^2(e)$:	9.8160E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	3.1331E+00

b) Tři uzlové body

Reziduální součet čtverců, RSC:	1.8054E+02
Průměr absolutních hodnot reziduí, Me:	2.0286E+00
Průměr relativních reziduí, Mer[%]:	40.830
Odhad reziduálního rozptylu, $s^2(e)$:	8.2063E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.8647E+00

c) Čtyři uzlové body

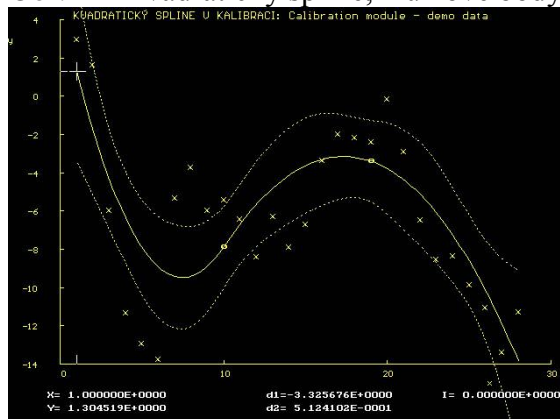
Reziduální součet čtverců, RSC:	7.9789E+01
Průměr absolutních hodnot reziduí, Me:	1.4011E+00
Průměr relativních reziduí, Mer[%]:	32.467
Odhad reziduálního rozptylu, $s^2(e)$:	3.7995E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	1.9492E+00

d) Pět uzlových bodů

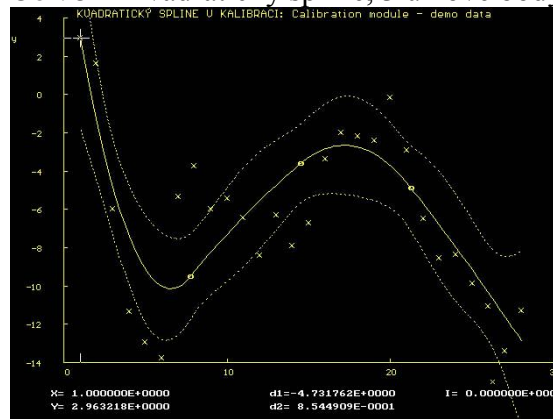
Reziduální součet čtverců, RSC:	9.2245E+01
---------------------------------	------------

Průměr absolutních hodnot reziduí, Me:	1.4185E+00
Průměr relativních reziduí, Mer[%]:	28.589
Odhad reziduálního rozptylu, $s^2(e)$:	4.6123E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.1476E+00
e) Šest uzlových bodů	
Reziduální součet čtverců, RSC:	9.7115E+01
Průměr absolutních hodnot reziduí, Me:	1.4633E+00
Průměr relativních reziduí, Mer[%]:	39.314
Odhad reziduálního rozptylu, $s^2(e)$:	5.1113E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.2608E+00

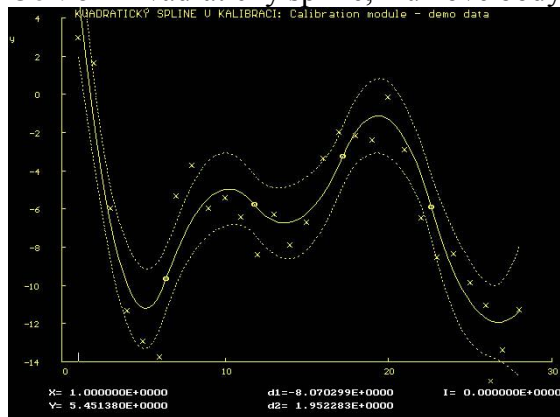
Obr. 4 – Kvadratický spline, 2 uzlové body



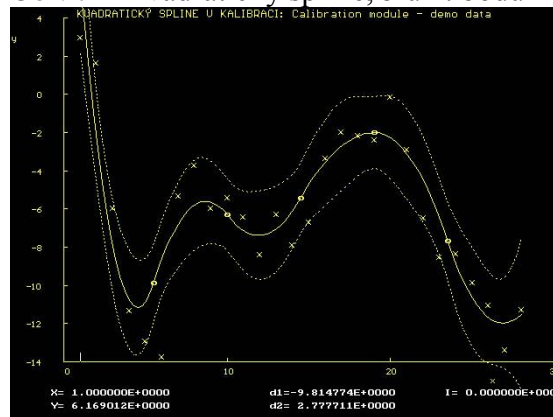
Obr. 5 – Kvadratický spline, 3 uzlové body



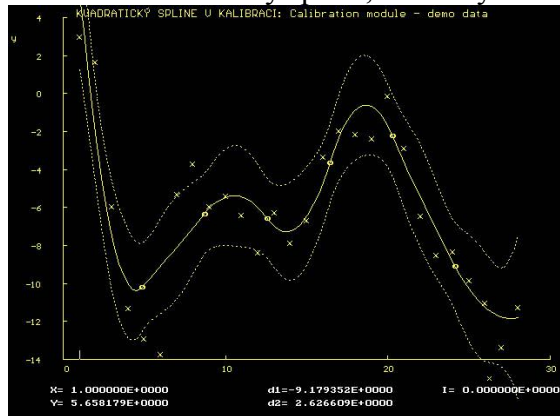
Obr. 6 – Kvadratický spline, 4 uzlové body



Obr. 7 – Kvadratický spline, 5 uzl. bodů



Obr. 7 – Kvadratický spline, 6 uzlových bodů



C) Kubický spline

a) Dva uzlové body

Reziduální součet čtverců, RSC:	1.7836E+02
Průměr absolutních hodnot reziduí, Me:	2.0233E+00
Průměr relativních reziduí, Mer[%]:	40.168
Odhad reziduálního rozptylu, $s^2(e)$:	8.1075E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.8474E+00

b) Tři uzlové body

Reziduální součet čtverců, RSC:	1.0476E+02
Průměr absolutních hodnot reziduí, Me:	1.5639E+00
Průměr relativních reziduí, Mer[%]:	31.159
Odhad reziduálního rozptylu, $s^2(e)$:	4.9885E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.2335E+00

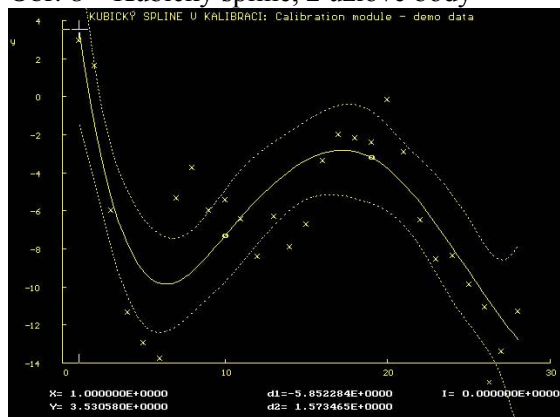
c) Čtyři uzlové body

Reziduální součet čtverců, RSC:	1.2837E+02
Průměr absolutních hodnot reziduí, Me:	1.6837E+00
Průměr relativních reziduí, Mer[%]:	31.483
Odhad reziduálního rozptylu, $s^2(e)$:	6.4187E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.5335E+00

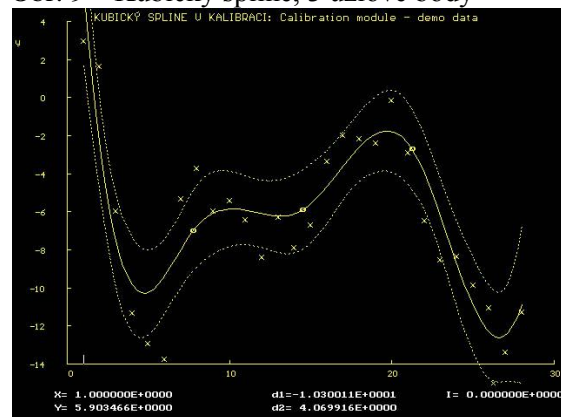
d) Pět uzlových bodů

Reziduální součet čtverců, RSC:	9.0038E+01
Průměr absolutních hodnot reziduí, Me:	1.4327E+00
Průměr relativních reziduí, Mer[%]:	46.787
Odhad reziduálního rozptylu, $s^2(e)$:	4.7389E+00
Odhad směrodatné odchylky reziduí, $s(e)$:	2.1769E+00

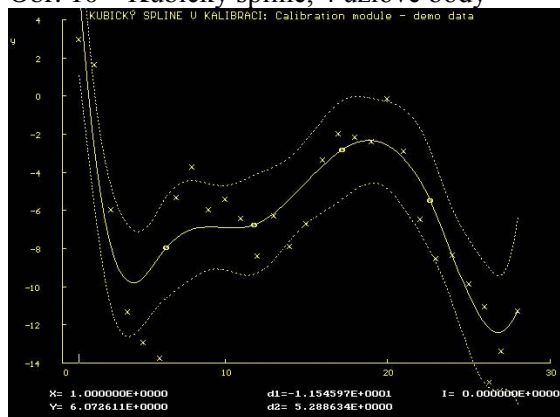
Obr. 8 – Kubický spline, 2 uzlové body



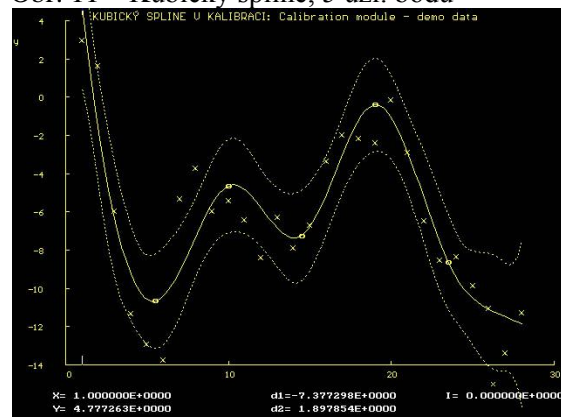
Obr. 9 – Kubický spline, 3 uzlové body



Obr. 10 – Kubický spline, 4 uzlové body



Obr. 11 – Kubický spline, 5 uzl. bodů



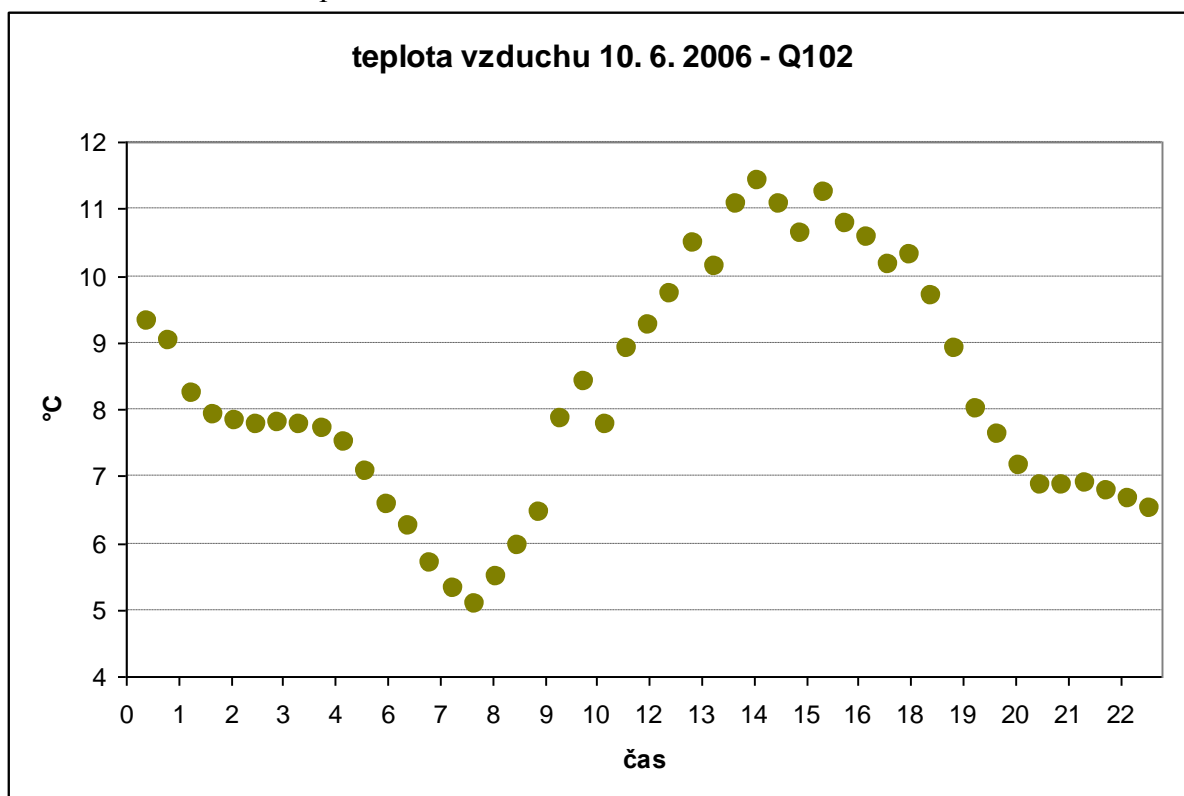
Závěr: Vyhodnocením grafů na obr. 2-11 a číselných kritérií provedené spline regrese byl jako nevhodnější model zvolen výsledek C^1 regrese s modelem ve tvaru kvadratické spline aproximace se 4 uzlovými body (bod B.c, obr. 6) s koeficienty rovnice $f[i]*x^2 + g[i]*x + h[i]$, pro $k[i-1] < x \leq k[i]$ v uzlových bodech $k[i]$:

	$k[i]$	$f[i]$	$g[i]$	$h[i]$
$k[1]$	6.4000E+00	9.7614E-01	-1.0023E+01	1.4498E+01
$k[2]$	1.1800E+01	-3.2448E-01	6.6254E+00	-3.8776E+01
$k[3]$	1.7200E+01	2.7770E-01	-7.5860E+00	4.5071E+01
$k[4]$	2.2600E+01	-4.5417E-01	1.7590E+01	-1.7144E+02
b	2.8000E+01	3.5649E-01	-1.9052E+01	2.4261E+02

Úloha 2: Vyhlazování experimentálních dat. Vstupní data jsou uvedena v souboru 4_4_data.xls, list „červen“. Jedná se o měření teploty vzduchu na výzkumné ploše Q102 dne 10. června 2006. Cílem je proložit křivku průběhem denních teplot měřených v deseti minutových intervalech, ze kterých jsou vypočteny půl hodinové průměry. Hlavním důvodem pro výpočet půl hodinových průměrů bylo snížení počtu řádků vstupní tabulky dat a tím pádem možnost zadat data do používané verze ADSTAT (max. 100 řádků).

Chceme, aby vyhlazující funkce byla dostatečně hladká a spojitá v první a druhé derivaci, tedy ze třídy C^2 .

Obr. 12 – Zobrazení vstupních dat



Kolísání dat v průběhu dne je způsobeno zejména měnící se oblačností nad měřícím bodem. Z tabulky dat je zřejmé, že se jedná o část dne, kdy je Slunce nad obzorem a teplota vzduchu v okolí čidla se mění při jasné a zatažené obloze.

Úloha byla řešena v programu ADSTAT 1.25 v modulu „Vyhlazování“.

Výstupy programu ADSTAT pro úlohu 2:

A) Reinsch spline

a) Parametr vyhlazení $S = 1$

Průměr residuí:	0.0000E+00
Průměr absolutních hodnot residuí:	9.5525E-02
Reziduální směrodatná odchylka:	1.4586E-01
Reziduální rozptyl:	2.1277E-02
Šikmost residuí:	7.3132E-01
Špičatost residuí:	4.8203E+00

b) Parametr vyhlazení $S = 1.5$

Průměr residuí:	-3.7795E-17
Průměr absolutních hodnot residuí:	1.1686E-01
Reziduální směrodatná odchylka:	1.7865E-01
Reziduální rozptyl:	3.1915E-02
Šikmost residuí:	5.7172E-01
Špičatost residuí:	4.6758E+00

c) Parametr residuí $S = 2$

Průměr residuí:	5.6692E-17
Průměr absolutních hodnot residuí:	1.4696E-01
Reziduální směrodatná odchylka:	2.0628E-01
Reziduální rozptyl:	4.2553E-02
Šikmost residuí:	1.6667E-01
Špičatost residuí:	4.0973E+00

d) Parametr residuí $S = 3$

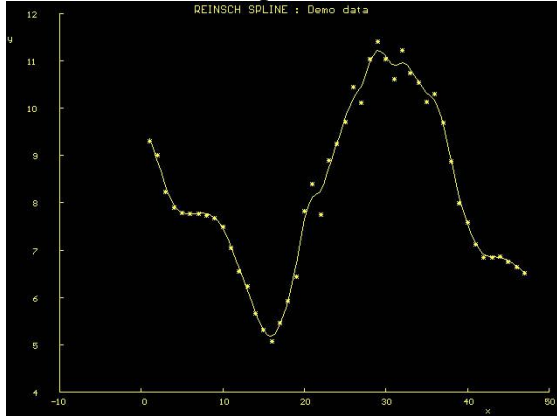
Průměr residuí:	1.8897E-17
Průměr absolutních hodnot residuí:	2.0080E-01
Reziduální směrodatná odchylka:	2.5265E-01
Reziduální rozptyl:	6.3830E-02
Šikmost residuí:	-1.9786E-01
Špičatost residuí:	3.0198E+00

e) Parametr vyhlazení $S = 4$

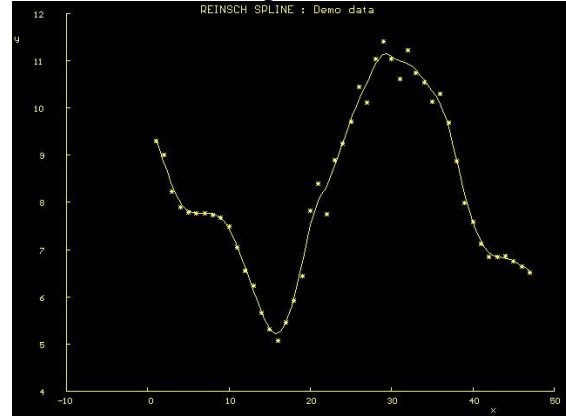
Průměr residuí:	7.5590E-17
Průměr absolutních hodnot residuí:	2.4065E-01
Reziduální směrodatná odchylka:	2.9173E-01
Reziduální rozptyl:	8.5106E-02
Šikmost residuí:	-1.5438E-01
Špičatost residuí:	2.4857E+00

Jako nejvhodnější bylo vybráno vyhlazení s parametrem $S = 2$ (obr. 15). Křivka leží v dostatečné blízkosti k experimentálně změřeným bodům a přitom je dostatečně hladká. Jako vhodné proložení se jeví také křivka na obr. 16 s parametrem vyhlazení $S = 3$, proložení experimentálních bodů je méně těsné.

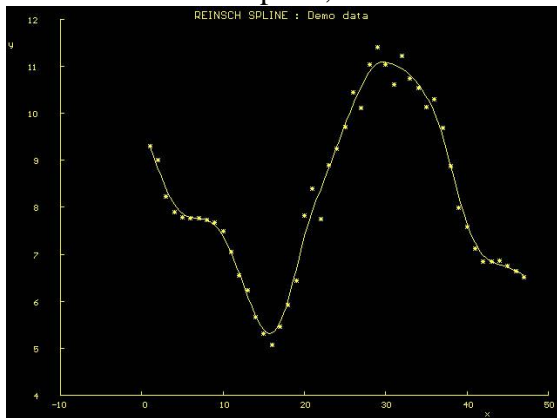
Obr. 13 – Reinsch spline, $S = 1$



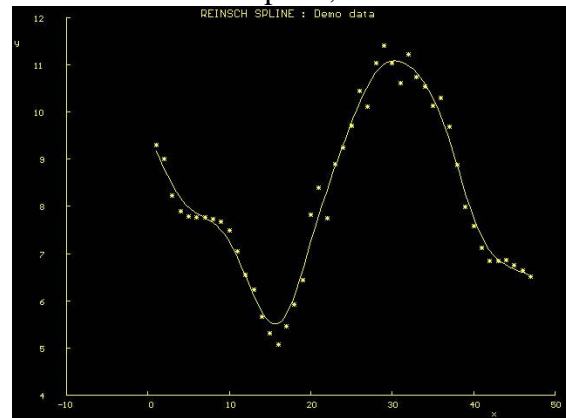
Obr. 14 – Reinsch spline, $S = 1.5$



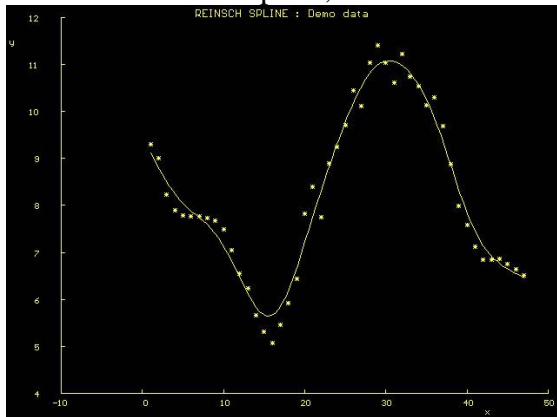
Obr. 15 – Reinsch spline, $S = 2$



Obr. 16 – Reinsch spline, $S = 3$



Obr. 17 – Reinsch spline, $S = 4$



B) Späth spline

a) Parametr vyhlazení $P = 0.1$

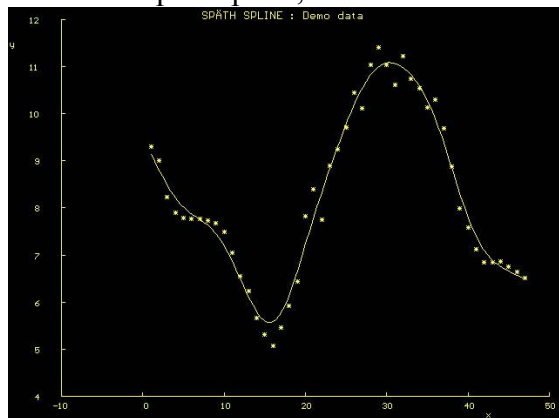
Průměr residuí:	1.8897E-17
Průměr absolutních hodnot residuí:	2.1984E-01
Reziduální směrodatná odchylka:	2.7065E-01
Reziduální rozptyl:	7.3251E-02
Šikmost residuí:	-2.0127E-01
Špičatost residuí:	2.7343E+00

b) Parametr vyhlazení $P = 1$

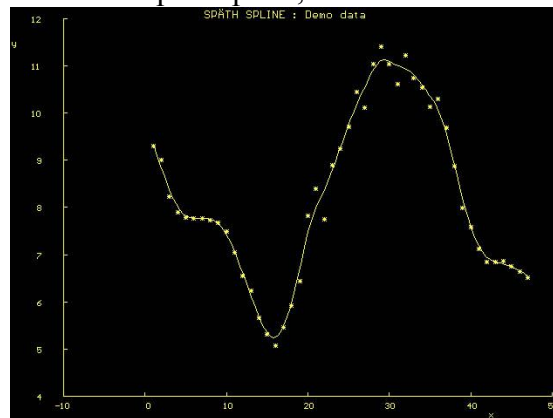
Průměr residuí:	-7.5590E-17
Průměr absolutních hodnot residuí:	1.2381E-01

Reziduální směrodatná odchylka:	1.8678E-01
Reziduální rozptyl:	3.4887E-02
Šikmost reziduí:	4.6787E-01
Špičatost reziduí:	4.5518E+00
b) Parametr vyhlazení $P = 3$	
Průměr reziduí:	0.0000E+00
Průměr absolutních hodnot reziduí:	1.0196E-01
Reziduální směrodatná odchylka:	1.5703E-01
Reziduální rozptyl:	2.4658E-02
Šikmost reziduí:	7.1573E-01
Špičatost reziduí:	4.8123E+00
b) Parametr vyhlazení $P = 5$	
Průměr reziduí:	7.5590E-17
Průměr absolutních hodnot reziduí:	9.2296E-02
Reziduální směrodatná odchylka:	1.4045E-01
Reziduální rozptyl:	1.9727E-02
Šikmost reziduí:	7.2951E-01
Špičatost reziduí:	4.8172E+00

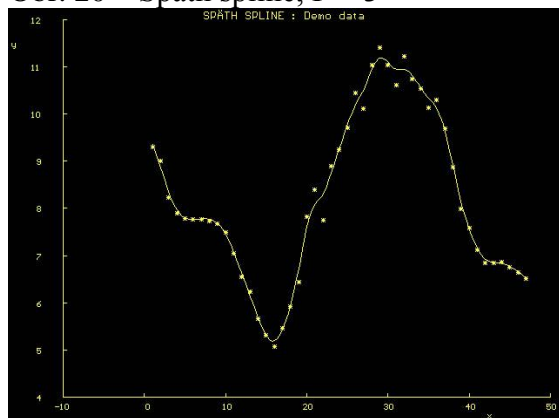
Obr. 18 – Späth spline, $P = 0.1$



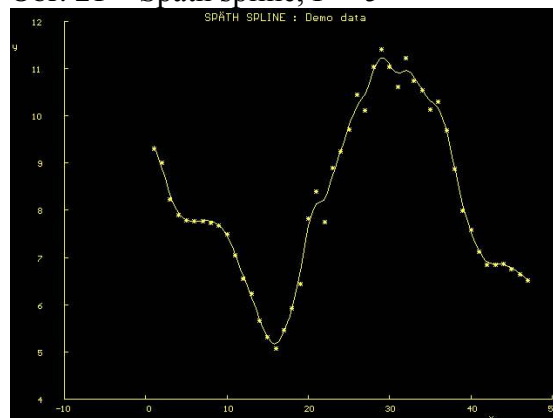
Obr. 19 – Späth spline, $P = 1$



Obr. 20 – Späth spline, $P = 3$



Obr. 21 – Späth spline, $P = 5$



Bylo dosaženo podobných výsledků jako u předchozí varianty. Křivka na obr. 19 je zcela srovnatelná s křivkou na obr. 14. Hladká křivka s parametrem vyhlazení $P = 0.1$ odpovídá výsledku dosaženému u Reinsch spline pro parametr vyhlazení $S = 4$.

C) Savitzki – Golay spline

a) Parametr vyhlazení $M = 5$

Průměr residuí:	7.5452E-04
Průměr absolutních hodnot residuí:	1.8946E-01
Reziduální směrodatná odchylka:	2.4786E-01
Reziduální rozptyl:	6.1434E-02
Šikmost residuí:	5.8908E-02
Špičatost residuí:	3.4767E+00

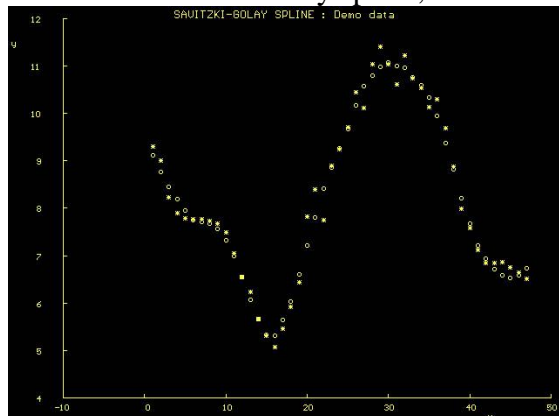
b) Parametr vyhlazení $M = 7$

Průměr residuí:	1.4389E-02
Průměr absolutních hodnot residuí:	2.6970E-01
Reziduální směrodatná odchylka:	3.2253E-01
Reziduální rozptyl:	1.0403E-01
Šikmost residuí:	-1.3402E-01
Špičatost residuí:	2.3393E+00

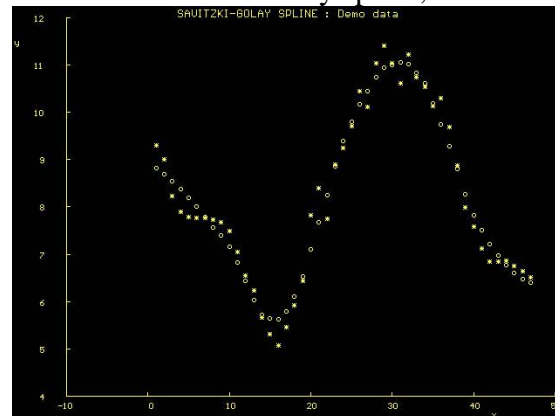
b) Parametr vyhlazení $M = 9$

Průměr residuí:	6.9801E-03
Průměr absolutních hodnot residuí:	3.5056E-01
Reziduální směrodatná odchylka:	4.2021E-01
Reziduální rozptyl:	1.7658E-01
Šikmost residuí:	7.6242E-02
Špičatost residuí:	2.1036E+00

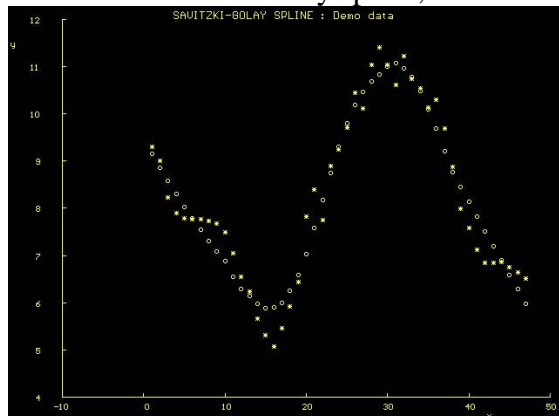
Obr. 22 – Savitzki-Golay spline, $M = 5$



Obr. 23 – Savitzki-Golay spline, $M = 7$



Obr. 24 – Savitzki-Golay spline, $M = 9$

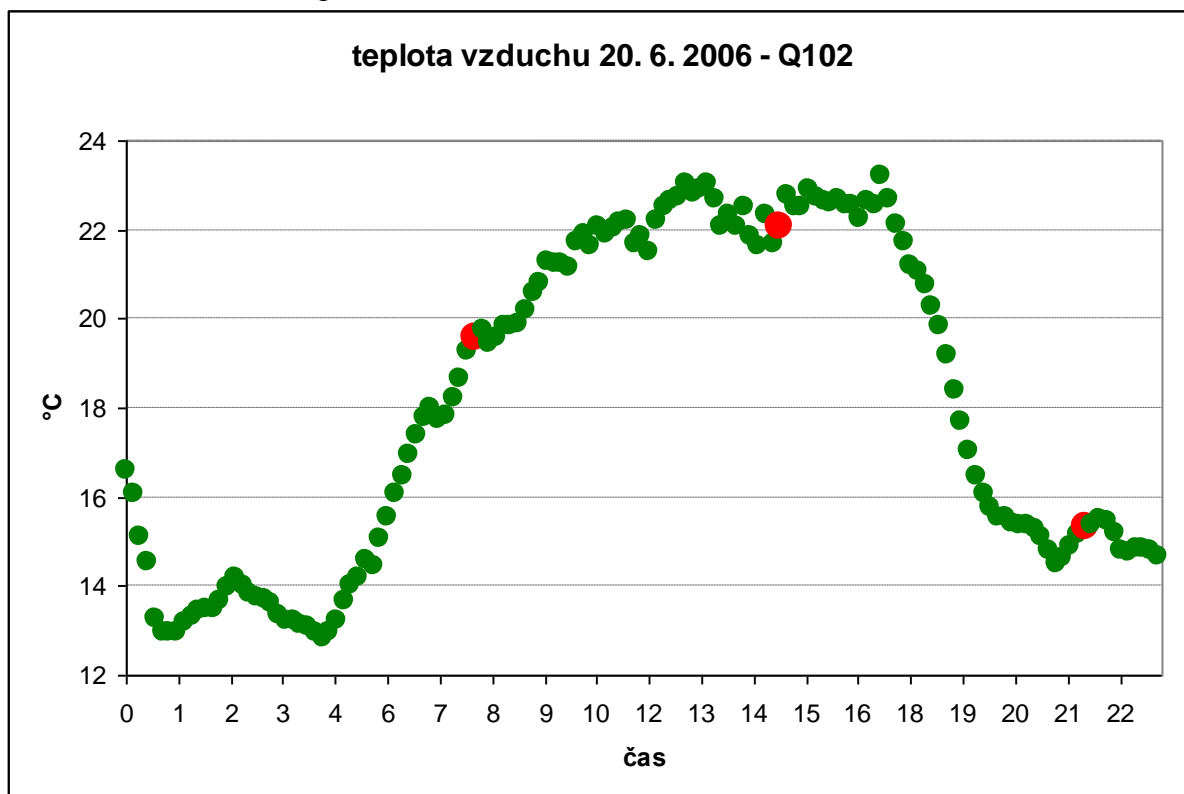


Využitím algoritmu Savitzki – Golay bylo nejlepšího výsledku dosaženo při parametru vyhlazení $M = 5$ (obr. 22).

Úloha 3: Interpolace. Cílem je proložit křivku naměřenými hodnotami teploty vzduchu z 20. 6. 2006 na výzkumné ploše Q102. Body, kterými má křivka procházet, jsou hodnoty naměřené v čase (SEČ) 7:00, 14:00, 21:00. Jde o hodnoty, které se používají pro výpočet průměrné denní teploty vzduchu.

Vstupní data jsou v souboru 4_4_data.xls, list „cerven2“. Pro možnost počítat v programu ADSTAT 1.25 byly spočteny půl hodinové průměry, aby došlo ke snížení počtu řádků tabulky ze 144 na nižší počet.

Obr. 25 – Zobrazení vstupních dat:



Úloha byla řešena v programu ADSTAT 1.25 v modulu „Kalibrace“. Použit byl kubický spline v podmínkách byla vybrána možnost „zadávání uzlů“. Zadány byly body, které odpovídají času (SEČ) 7:00, 14:00 a 21:00.

Nalezení koeficientů rovnice bylo pro srovnání provedeno i pro variantu „konstantní uzlové intervaly“, kdy byly zvoleny tři uzlové body.

Výstupy programu ADSTAT pro úlohu 3:

A) Zadané uzly

Hodnoty uzlů:

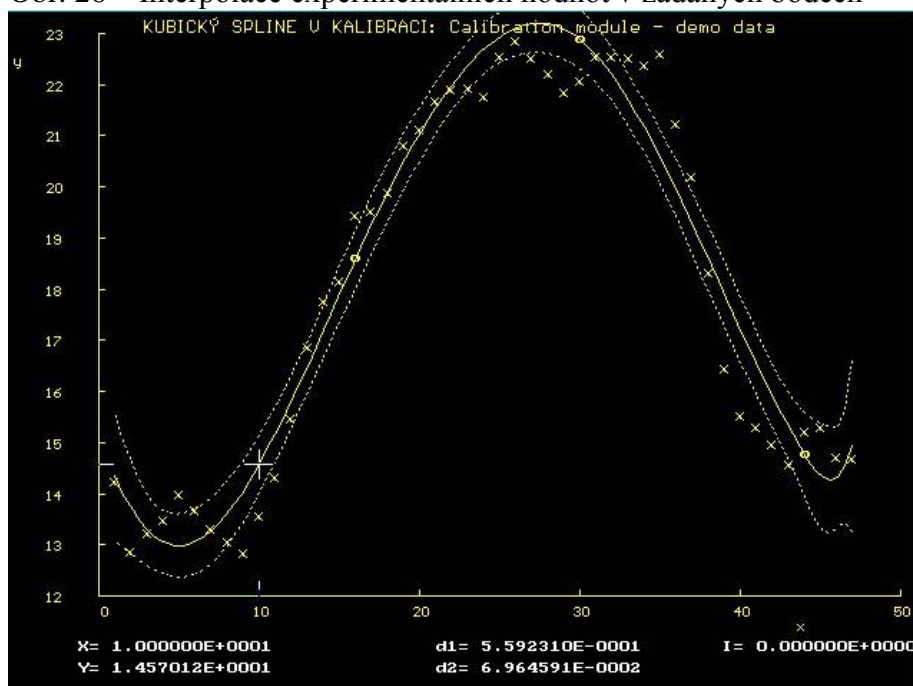
- k[1]: 1.6000E+01 (7:00 SEČ)
- k[2]: 3.0000E+01 (14:00 SEČ)
- k[3]: 4.4000E+01 (21:00 SEČ)

Koeficienty rovnice: $e[i]*x^3 + f[i]*x^2 + g[i]*x + h[i]$, pro $k[i-1] < x \leq k[i]$

	k[i]	e[i]	f[i]	g[i]	h[i]
k[1]	1.6000E+01	-2.7179E-03	1.1636E-01	-9.5259E-01	1.5178E+01
k[2]	3.0000E+01	-9.1540E-04	2.9840E-02	4.3172E-01	7.7950E+00
k[3]	4.4000E+01	2.0679E-03	-2.3866E-01	8.4866E+00	-7.2754E+01
b	4.7000E+01	5.1346E-02	-6.7433E+00	2.9469E+02	-4.2704E+03

Reziduální součet čtverců, RSC: 2.9858E+01
 Průměr absolutních hodnot reziduí, Me: 6.4942E-01
 Průměr relativních reziduí, Mer[%]: 3.712
 Odhad reziduálního rozptylu, $s^2(e)$: 7.4646E-01
 Odhad směrodatné odchylky reziduí, $s(e)$: 8.6398E-01

Obr. 26 – Interpolace experimentálních hodnot v zadaných bodech



B) Konstantní uzlové intervaly

Hodnoty uzlů:

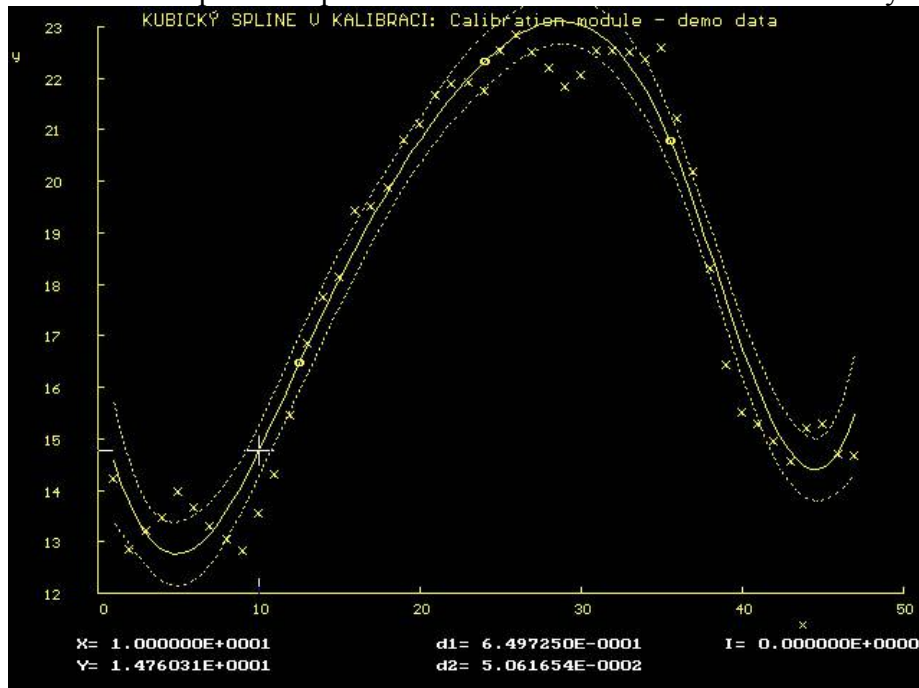
k[1]: 1.2500E+01
 k[2]: 2.4000E+01
 k[3]: 3.5500E+01

Koeficienty rovnice: $e[i]*x^3 + f[i]*x^2 + g[i]*x + h[i]$, pro $k[i-1] < x \leq k[i]$

	k[i]	e[i]	f[i]	g[i]	h[i]
k[1]	1.2500E+01	-4.9521E-03	1.7387E-01	-1.3421E+00	1.5746E+01
k[2]	2.4000E+01	-2.8347E-04	-1.2020E-03	8.4634E-01	6.6276E+00
k[3]	3.5500E+01	-1.3879E-03	7.8318E-02	-1.0621E+00	2.1895E+01
b	4.7000E+01	8.1980E-03	-9.4258E-01	3.5180E+01	-4.0697E+02

Reziduální součet čtverců, RSC: 2.2591E+01
 Průměr absolutních hodnot reziduí, Me: 5.6794E-01
 Průměr relativních reziduí, Mer[%]: 3.341
 Odhad reziduálního rozptylu, $s^2(e)$: 5.6477E-01
 Odhad směrodatné odchylky reziduí, $s(e)$: 7.5151E-01

Obr. 27 – Interpolace experimentálních hodnot s konstantními uzlovými intervaly



Závěr: Bylo nalezeno proložení experimentálních bodů pevně stanovenými uzlovými body (interpolace) zvolenými v časech používaných pro výpočet průměrné denní teploty vzduchu. Při řešení úlohy s konstantními uzlovými intervaly (aproximace) bylo dosaženo lepšího proložení bodů křivkou. Uzlové body se nacházejí v časech měření mezi 6:00 a 6:30, 12:00 a mezi 17:30 a 18:00.