

UNIVERZITA TOMÁŠE BATI VE ZLÍNE
Fakulta managementu a ekonomiky

DIPLOMOVÁ PRÁCE

2005

Roman KASAL

Projekt užití statistických metod v procesu plánování výroby ve firmě Mitas, a.s.

Bc. Roman Kasal



Univerzita Tomáše Bati ve Zlíně
Fakulta managementu a ekonomiky
Ústav podnikové ekonomiky
akademický rok: 2004/2005

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Roman KASAL**
Studijní program: **N 6208 Ekonomika a management**
Studijní obor: **Podniková ekonomika**

Téma práce: **Projekt užití statistických metod v procesu plánování výroby ve firmě Mitas, a.s.**

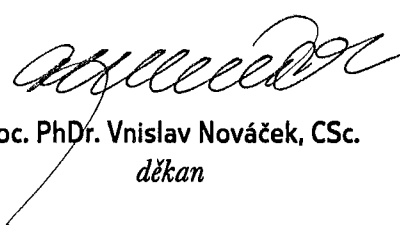
Zásady pro vypracování:

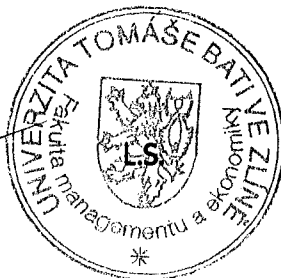
- 1. Vymezte rámec aktuálních projektových příležitostí firmy Mitas, a.s., definujte ideu a hypotézu projektového řešení.**
- 2. Konfrontujte záměr ideje a hypotézy, dostupné literární prameny z oblastí statistických metod v procesu plánování výroby spolu se zpracovanou analýzou k podpoře projektového řešení.**
- 3. Zpracujte syntézu poznatků vyplývajících z bodu 2.**
- 4. Formulujte diagnózu projektu jako určující východisko pro adresné zpracování projektu.**
- 5. Vypracujte projekt užití statistických metod v procesu plánování výroby.**
- 6. Verifikujte projektové řešení v návaznosti na bod 5.**
- 7. Prezентуйте a zdůvodněte efektivnost daného projektového řešením spolu se závěrečným doporučením pro společnost Mitas, a.s.**

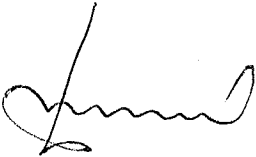
Rozsah práce: **70 stran**
Rozsah příloh:
Forma zpracování diplomové práce: **tištěná/elektronická**
Seznam odborné literatury:
Dle doporučení vedoucího diplomové práce.

Vedoucí diplomové práce: **Ing. Petr Klímek, Ph.D.**
Ústav informatiky a statistiky
Datum zadání diplomové práce: **7. března 2005**
Termín odevzdání diplomové práce: **13. května 2005**

Ve Zlíně dne 21. února 2005


doc. PhDr. Vnislav Nováček, CSc.
děkan




prof. Ing. Jiří Polách, CSc.
ředitel ústavu

RESUMÉ

Anotace česky

Diplomová práce se ujímá snahy o přiblížení aplikace užitečných statistických i nestatistických metod k popisu a ovládnutí reálných skutečností v procesu plánování výroby motoplášťů firmy Mitas, a.s. Více než sedmdesáti stránková snaha čistého textu využívá statistických a nestatistických nástrojů týkajících se především časových řad, statistické indukce a genetických algoritmů k návrhu řešení optimalizace.

Anotace ve světovém jazyce

My diploma's work is trying to provide an access to utilization of useful statistics and non statistic methods and algorithms in planning process in sphere of motorcycle tires business in Mitas join-stock company. Work applies and emphasizes statistics tools for time series, descriptive statistics and genetic algorithms designed to optimalization.

Na tomto místě bych chtěl poděkovat svému vedoucímu diplomové práce ing. Petru Klímkovi Ph.D. za jeho více než kladný přístup a užitečnou pomoc a svému konzultantovi Josefu Křemečkovi. Dále chci poděkovat pracovníkům firmy Mitas, a.s. Zlín, kteří mi umožnili vypracovat tento diplomový projekt, a především Bc. Martinu Juráskovi, pracovníkovi výše zmíněné firmy, bez jehož “propagační akce“ ve firmě Mitas, a.s. bych tuto příležitost jistě nedostal.

OBSAH

OBSAH	5
ÚVOD	7
I. TEORETICKÁ ČÁST	9
1 VÝBĚR STATISTICKÝCH A NESTATISTICKÝCH METOD A ALGORITMŮ PRO PROCES PLÁNOVÁNÍ VÝROBY	10
1.1 ČASOVÉ ŘADY	10
1.1.1 Metoda klouzavých průměrů.....	10
1.1.2 Regresní analýza	11
1.1.3 Aditivní modely pro analýzu sezónní složky	12
1.1.4 Harmonická analýza.....	13
1.1.5 SARIMA modelování	15
1.1.6 Umělé neuronové sítě	18
1.1.7 Analýza reziduí	20
1.1.8 M.S.E. statistika	21
1.2 STATISTICKÁ INDUKCE	21
1.2.1 Odhad.....	22
1.2.2 Statistická hypotéza	22
1.3 GENETICKÉ ALGORITMY	25
1.3.1 Princip	25
1.3.2 Problém obchodního cestujícího	26
II. ANALYTICKÁ ČÁST	30
2 ANALÝZA SOUČASNÉHO STAVU A ZÁKLADNÍ ANALÝZA DAT	31
2.1 SPOLEČNOST MITAS, A.S.	31
2.2 VÝROBA MOTOPLÁŠŤŮ	32
2.3 VYHODNOCENÍ SOUČASNÉHO STAVU APLIKACE STATISTICKÝCH METOD	34
2.4 MATLAB	34
2.5 ANALÝZA DAT ZÁKLADNÍMI STATISTICKÝMI NÁSTROJI	35
2.5.1 Očištění dat o nepravidelnosti	36
2.5.2 Očištění dat o sezónní výkyvy	37
2.5.3 Nalezení trendu očištěných dat.....	38
2.5.4 Nalezení trendu prodejů v závislosti na pevných objednávkách nástrojem lineární regrese.....	40
2.5.5 Nalezení trendu prodejů modelem konstantní sezónnosti s lineárním trendem	41

III. PRAKTICKÁ ČÁST.....	43
3 APLIKACE POKROČILÝCH STATISTICKÝCH A NESTATISTICKÝCH METOD A ALGORITMŮ	44
3.1 ČASOVÉ ŘADY	44
3.1.1 Aplikace harmonické analýzy	44
3.1.2 Komplettní regresní model pro sezónní složku.....	49
3.1.3 Durbin -Watsonova analýza náhodné složky.....	50
3.1.4 Výběr nejvhodnějšího modelu, různé statistiky, predikce a konečná úprava dat.....	52
3.1.5 Box-Jenkinsova metodologie	53
3.1.6 Umělé neuronové sítě a jejich využití pro časové řady	60
3.2 STATISTICKÉ NÁSTROJE PRO POPIS VÝROBY	66
3.2.1 Volba vhodných statistických nástrojů.....	66
3.3 APLIKACE GENETICKÝCH ALGORITMŮ V PROCESU PLÁNOVÁNÍ VÝROBY.....	70
3.3.1 Rozplánování výroby v úzkém místě provozu.....	70
3.3.2 Problém obchodního cestujícího	76
3.4 IMPLEMENTACE STATISTICKÝCH METOD DO PROCESU PLÁNOVÁNÍ VÝROBY	81
ZÁVĚR.....	82
SEZNAM POUŽITÉ LITERATURY.....	84
SEZNAM OBRÁZKŮ.....	87
SEZNAM TABULEK	89
SEZNAM PŘÍLOH.....	90

ÚVOD

Problém nepříliš častého využití statistických metod není nijak výjimečný. I v této práci se setkáváme se zmíněným neduhem, přestože je aplikace mnohdy jediným spolehlivým popisem skutečnosti. Zmíněný nedostatek má své kořeny již v období studia více či méně ekonomického charakteru, kde by měl být kladen důraz na ovládání alespoň základních, velmi užitečných metod. V drtivé většině se přístup k učivu vyskytuje jen ve dvou formách. První je přístup relativně nekompromisního docenta statistiky, který vyžaduje rychlé strávení ne snadno stravitelného učiva, což vyúsťuje k problémům ve studiu, druhý pak přístup naprosto opačný. Bylo by dobré alternativy sloučit a přidat důraz především na spolehlivé a ověřené ovládnutí základních metod a schopnost pracovat s literaturou našeho tématu. Řešení však v této práci po mně není vyžadováno a nic není tak jednoduché, jak se zdá. Negativum mé práce ovšem leží na stejném základě: neochota přistoupit ke složitě vypadajícím vzorcům (kterou se snažím odstranit alespoň k sobě samému). Chybovost v aplikaci statistických nástrojů bez důraznějšího vedení v učivu je téměř neodstranitelná. „Učí-li se člověk sám, je sdělovací a zpracovávající síla totožná a pokroky musí být menší a pomalejší,“ jak velmi výstižně a pravdivě popisuje Goethe.

Cíl položeného úkolu tkví v návrhu aplikace statistických metod v běžném procesu plánování výroby motoplášťů od prvotní analýzy prodeje ke konečnému rozplánování výroby. Diplomová práce poskytuje konkrétní řešení v kapitolách 2.5, 3.1, 3.2 a 3.3.

První kapitola, základní analýza dat, poskytuje doporučení zejména v popisu a přípravě dat pro další zpracování.

Druhá kapitola se týká oblasti časových řad. Statistických nástrojů prakticky není mnoho, pouze klouzavé průměry, regresní modely s případnými různými aditivními modifikacemi, harmonická analýza, SARIMA modelování a umělé neuronové sítě.

Třetí kapitola je probrána velmi zběžně z důvodu omezení rozsahu práce, přesto logicky nevyhnutelná. Zaměřuje se na jen nejzákladnější popisné metody, které považují za velmi důležité. Další studium pro schopnost ovládnutí popisné statistiky více než doporučuji.

Čtvrtá kapitola vychází z předchozích kapitol, využívá jejich výsledky. Obsahuje základ nejmodernějšího a nejúspěšnějšího přístupu k nalezení extrému složitější funkce, tedy genetické algoritmy. Aplikace stochastických metod je neustále nezbytná z příčiny zatím chabého

dostupného výpočetního výkonu. Genetické algoritmy mohou být aplikovány kdekoli je třeba minimalizovat či maximalizovat (optimalizovat) jakkoliv vytvořenou funkci, například funkci nákladů či výnosů.

I. TEORETICKÁ ČÁST

1 VÝBĚR STATISTICKÝCH A NESTATISTICKÝCH METOD A ALGORITMŮ PRO PROCES PLÁNOVÁNÍ VÝROBY

Přestože se na nižším místě neobjeví citace z [7], považuji za nezbytné literaturu zmínit. Důvod leží v panu Dostálovi, který nám odstranil “strašáka” bránící přístup k moderním metodám analýz dat v rámci předčasně zrušeného oboru. Tato jednoduše pojatá skripta mi otevřela cestu k daleko složitějšímu a hlubšímu pochopení následujících řádek a koneckonců k pojetí mé diplomové práce.

1.1 Časové řady

1.1.1 Metoda klouzavých průměrů

Metoda klouzavých průměrů se řadí mezi adaptivní přístupy k trendové složce. Adaptivní přístupy lze charakterizovat tak, že jsou schopny pracovat s trendovými složkami, které mění v čase globálně svůj charakter, takže pro jejich popis nelze použít žádnou matematickou křivku s neměnnými parametry. Na druhé straně se však předpokládá, že v krátkých úsecích časové řady je takové vyrovnaní pomocí matematických křivek možné, i když tyto křivky mají obvykle v různých úsecích odlišné parametry. Je tedy možné pouze lokální vyrovnaní trendu, takže se někdy mluví o tzv. koncepci postupného trendu.

Předpokládejme, například, že chceme polynomem třetího řádu vyrovnat lichý počet hodnot uvažované řady, tedy $y_{t+t}; t=1,2,\dots,n; t=\frac{t+1}{2}-(n-1)$, pak koeficienty vyrovnávajícího polynomu o řádu j určíme metodou nejmenších čtverců tak, že minimalizujeme výraz

$$\sum_{t=\min}^{t=\max} (y_{t+t} - b_0 - b_1 t^1 - \dots - b_j t^j). \quad (1)$$

Nás přitom zajímá pouze odhad b_0 , neboť je to hodnota vyrovnávajícího polynomu $b_0 + b_1 t^1 + \dots + b_j t^j$.

Například pro $j=3$ platí $b_0 = \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2})$, takže odhadnutá trendová složka a zároveň vyrovnaná hodnota řady v čase t je rovna $\hat{y}_t = b_0$.

Pro sudý počet hodnot má pak funkce tvar pro předpoklad, že $\hat{y}_t = b_0$ a

$$\bar{y}_t = \frac{(\bar{y}_t + \bar{y}_{t+1})}{2}, \quad (2)$$

který nazýváme centrovaný klouzavý průměr [3].

1.1.2 Regresní analýza

Regresní analýza slouží k poznání a matematickému popisu jednostranných statistických závislostí. Jde o situaci, kdy proti sobě stojí vysvětlující (nezávislé) proměnné a vysvětlované (závislé) proměnné. Odhadem je pak funkce $\hat{y} = f(x; b_0, b_1, \dots, b_p)$, kdy b jsou parametry funkce nezávislé proměnné x . Regresní analýza je založena na minimalizaci odchylek zjištěných hodnot od proložené regresní funkce, tzn. $\sum_{i=1}^n (y_i - Y)^2 = \min$, kde y_i je naměřená hodnota a Y je námi hledaná funkce. Pak minimalizaci provedeme parciální derivací, podle parametrů funkce, této rovnice položené nule a vypočteme v soustavě normálních rovnic jednotlivé parametry.

Popřípadě můžeme použít maticovou metodu, kde:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \text{M} \\ y_n \end{pmatrix}, \quad F = \begin{pmatrix} 1 & f_{11} & \mathbf{K} & f_{1p} \\ 1 & f_{21} & \mathbf{L} & f_{2p} \\ \text{M} & \text{M} & \mathbf{O} & \text{M} \\ 1 & f_{n1} & \mathbf{L} & f_{np} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \text{M} \\ b_p \end{pmatrix}$$

pak $b = (F'F)^{-1}F'y$.

Pro posouzení spolehlivosti koeficientů použijeme t test, kde daný parametr přijmeme, když

$$t = \frac{b}{SE_b} > t_{0,05}(n - k - 1), \text{ kde } k \text{ je počet parametrů.}$$

Výpočet směrodatné odchylka b má tvar

$$SE_b = \frac{s_y}{\sqrt{ns_x^2}}, \quad (3)$$

kde s_x^2 je prostý rozptyl hodnot X .

Interval spolehlivosti koeficientu b , což je skutečný, neznámý parametr, je dán s 95% pravděpodobností výrazem $b = b \pm t_{0,025}(n-k-1)SE_b$. Při $n > 30$ lze t -statistiku aproximovat na normální rozdělení.

Pro zjištění vhodného odhadu regresní roviny provedeme celkový F -test,

$$\text{kde } F = \frac{\frac{\sum (Y_i - \bar{y})^2}{k-1}}{\frac{\sum (y_i - Y_i)^2}{n-k-1}} > F_{1-\alpha}(k; n-k-1), \text{ kde } k \text{ je počet koeficientů } b.$$

Zlomek představuje: $F = \frac{\text{vysvětlěný rozptyl}}{\text{nevysvětlěný rozptyl}}$, blíže [11], [12], [22], [23] a [27], .

1.1.3 Aditivní modely pro analýzu sezónní složky

Při analýze časových řad s periodou kratší než jeden rok se setkáváme téměř vždy s existencí sezónních vlivů reprezentovaných v časových řadách sezónní složkou. Protože sezónní kolísání do značné míry zakrývá dynamiku ekonomických jevů, provádí se tzv. sezónní očišťování, jehož úkolem je vyloučit sezónní složku z časové řady. Úkolem popisu sezónní složky je tedy jednak kvantifikace sezónní složky za účelem analýzy sezónnosti a jednak výpočet sezónně očištěné časové řady [13].

Model konstantní sezónnosti

U tohoto modelu uvažujeme pravidelné opakování sezónních vlivů. Sezónní výkyvy se pravidelně realizují v každém dílčím období j v určité absolutní výši.

Model se dále dělí na *model konstantní sezónnosti se schodovitým trendem* a *model konstantní sezónnosti s ročním lineárním trendem*.

Model proporcionální sezónnosti

Tento model zavádíme, když není splněn předpoklad konstantní sezónnosti, tj. v období j se sezónní výkyvy mění přímo úměrně dosažené úrovni trendové složky.

Model smíšené sezónnosti

Model zahrnuje prvky modelů konstantní a proporcionální sezónnosti. Předpokládáme, že sezónní složka se skládá ze dvou částí, a to *konstantního sezónního výkyvu* a *proměnlivé části, která se mění se stejnou intenzitou*.

1.1.4 Harmonická analýza

Protože data mohou vykazovat jak cyklickou, tak sezónní složku, budeme raději hovořit o periodické složce. Vedle klasického přístupu však lze vycházet i z jiné koncepce, kdy předpokládáme, že časovou řadu možno popsat ‘směsí’ sinusovek a kosinusovek o rozličných amplitudách a frekvencích. Tato koncepce umožní provést explicitní popis periodického chování časové řady a především vystopovat ty významné složky periodicity, které se podílí na věcných vlastnostech zkoumaného procesu. Uvažujme řady s měnícím se trendem.

Časovou řadu definujme ve tvaru trigonometrické řady jako

$$Y_t = T_t + \sum_{j=1}^H a_j \sin(w_j t) + \sum_{j=1}^H b_j \cos(w_j t) + e, \quad (4)$$

kde $H = n/2$ pro sudé či $H = (n-1)/2$ pro n liché, $w_j = 2\pi j/n$ pro $j = 1, \dots, H$ je j -tá frekvence a a_j a b_j jsou koeficienty, jejichž hodnoty musíme teprve určit. T_t může představovat být libovolnou trendovou funkci.

Reální hodnoty je nejprve nutné očistit od trendu $\tilde{Y}_t = Y_t - T_t$, pak přistoupíme k výpočtu koeficientů metodou nejmenších čtverců, minimalizujeme výraz

$$\sum_{t=1}^n \left(\tilde{Y}_t - m_0 - \sum_{j=1}^H a_j \sin(w_j t) - \sum_{j=1}^H b_j \cos(w_j t) \right)^2, \quad (5)$$

pochopitelně $m_0 = 0$, pak koeficienty se rovnají

$$a_j = \frac{2}{n} \sum_{t=1}^n \tilde{Y}_t \sin(w_j t) \quad b_j = \frac{2}{n} \sum_{t=1}^n \tilde{Y}_t \cos(w_j t) \quad (6)$$

Periodogram

Periodogramem, viz Tab. 1., rozumíme soupis všech hodnot teoretických rozptylů vyjádřených pomocí amplitud kmitu, kde

$$\text{var}(\hat{Y}_t) = \frac{1}{n} \sum_{t=1}^n (\hat{Y}_t - \bar{y}_t)^2 = \frac{1}{n} \sum_{t=1}^n (a_j^2 + b_j^2). \quad (7)$$

Tab. 1. Periodogram.

j	Frekvence $w_j(\text{rad})$	Fourierova perioda (měsíce)	periodogram
1	$p \cdot 1/12$	24	$\sum_{t=1}^n (a_j^2 + b_j^2) / n$
2	$p \cdot 2/12$	12	
...	...	...	
H	p	2	
Σ	X	x	$\text{var}(\bar{y})$

Fisherův test

Jeho testové kritérium je založeno na hodnotách periodogramu určených pro frekvence w_j . Testuje se nulová hypotéza H_0 , že $y_1, \dots, y_n$ je posloupnost nezávislých náhodných veličin, majících normální rozdělení.

V testu postupujeme tak, že vypočítané hodnoty periodogramu seřadíme sestupně podle velikosti a přečíslováme. Označíme n_1 jako největší z nich a n_H jako nejmenší z nich. Postupně se pak testuje, že

$$P\{W_j > g_f(a)\} \approx H[1 - g_f(a)]^{H-1}, \quad (8)$$

pro $H \leq 50$, kde $W_j = \frac{n_1}{\sum_{j=1}^H n_j}$ a $0 < g_f(a) < 1$. Pokud $P < a$, nulovou hypotézu zamítneme a

považujeme periodu za prokázanou.

Sieglův test

I když Fisherův test periodicity je nejvydatnějším postupem zmapování významu jednotlivých period, uvádí se, že jeho síla v situacích, kdy v časové řadě vedle sebe existují dvě i více významných period, klesá. Proto se v poslední době objevily jeho modifikace. Ty využívají výborných vlastností Fisherova testu při jednoduché periodicitě, ale rozvíjí je pro periodicitu složenou, která v realitě nemusí být zas tak výjimečná.

A. F. Siegel navrhuje použít místo statistiky W testové kritérium T_s , pak pokud

$$T_s = \sum_{j=1}^H [W_j - 0,6 \cdot g_f(a)] > 0 \quad (9)$$

kde tentokrát $W_j = \frac{n_1}{\sum_{j=1}^H n_j}$, přijímáme hypotézu o vlivu periody [2].

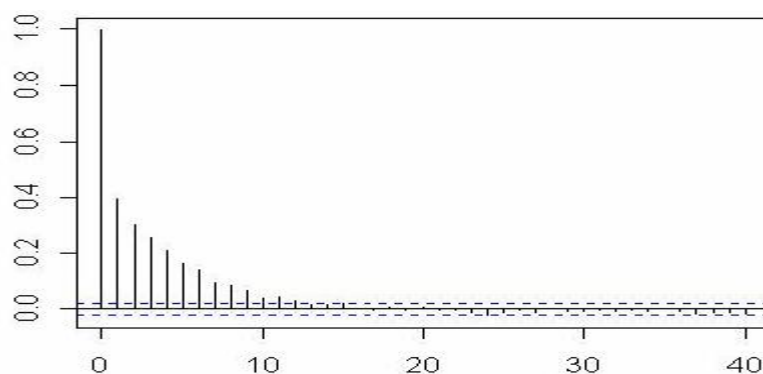
1.1.5 SARIMA modelování

ACF

V případě stacionárního stochastického procesu [1] lze vyjádřit každou autokorelační funkci jako podíl kovariance a korelace

$$r_k = \frac{E(X_t - m)(X_{t-k} - m)}{\sqrt{E(X_t - m)^2} \sqrt{E(X_{t-k} - m)^2}} \quad (10)$$

pak ACF, korelogram (viz Obr. 1.), graficky vyjadřuje korelační závislost X_t na X_{t-k} .



Obr. 1. ACF.

PACF

Korelace mezi dvěma náhodnými veličinami je často způsobena tím, že obě tyto veličiny jsou korelovány s veličinou třetí. Velká část korelace mezi veličinami X_t a X_{t-k} může být tedy způsobena jejich korelací s veličinami mezi hodnotami. Parciální autokorelace podávají informaci o korelaci veličin X_t a X_{t-k} očištěné o vliv veličin ležících mezi nimi.

Parciální autokorelaci se zpožděním k vyjadřuje parciální regresní koeficient f_{kk} v autoregresi k -tého řádu

$$X_t = f_{k1}X_{t-1} + f_{k2}X_{t-2} + \dots + f_{kk}X_{t-k} \quad (11)$$

Korelogram PACF je s ACF velmi podobný, viz Obr. 1.

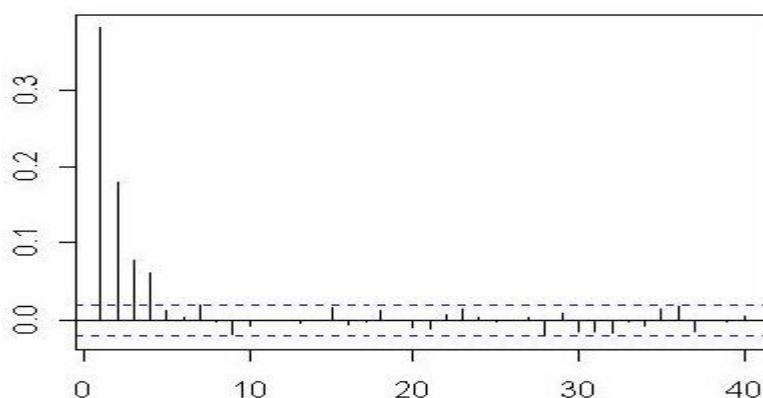
AR

Model AR řádu n se značí $AR(p)$ a je vyjádřen ve formě

$$X_t = f_1X_{t-1} + f_2X_{t-2} + \dots + f_nX_{t-p} + a_t, \quad (12)$$

a_t je bílý šum.

Model se dá odhalit grafickým zhodnocením korelogramů ACF a PACF, kde hodnoty o 1 až p hodnot zpět u ACF exponenciálně klesají, zatímco u PACF jsou prokázány právě a jen parciální autokorelace (viz Obr. 2.) do p -tého řádu a ostatní jsou konstantně nízké.



Obr. 2. PACF.

MA

Každý stacionární proces, který neobsahuje deterministickou složku, může být vyjádřen jako lineární kombinace řady nekorelovaných stejně rozdělených náhodných veličin. Tato lineární kombinace se označuje jako lineární proces, který lze vyjádřit jako

$$X_t - m = a_t + y_1 a_{t-1} + y_2 a_{t-2} + \dots, \quad (13)$$

kde $m = D(X_t)$.

Modely MA vycházejí z lineárního procesu (13), liší se pouze tím, že mají končený počet vahy.

Proces $MA(q)$ se vyznačuje stejnými vlastnostmi jako $AR(p)$, jen jsou úlohy ACF a PACF obrácené.

Stacionarizace

Abychom postoupili k dalším úpravám modelů ARMA, je nutné hodnoty nejprve stacionarizovat, a to v průměru a rozptylu.

V případě nutnosti stacionarizovat v průměru vykazuje korelogram ACF v ideálním případě tendenci pomalého klesání a PACF jen jednu prokázanou hodnotu. V případě lineárního trendu je třeba řadu diferencovat prvním řádem, v případě kvadratického trendu řádem druhým (podrobněji [1]). Jistě jsou povoleny i jiné trendy pro stabilizaci v průměru.

Otestovat řady bychom měli také v rozptylu. K tomu nám pomáhá Box-Coxova koncepce, která vychází ze vztahu

$$H(X_t) = X_t^{(l)} = \frac{X_t^l - 1}{l} \text{ pro } l \neq 0 \text{ a } H(X_t) = \ln(X_t) \text{ pro } l = 0, \quad (14)$$

kde l je transformační parametr. Tento parametr se nesnadno odhaduje, proto se používá následující tabulka.

Tab. 2. Hodnoty l s odpovídajícími transformacemi [1].

l	transformace
-1,0	$1/X_t$
-0,5	$1/\sqrt{X_t}$
0,0	$\ln(X_t)$
0,5	$\sqrt{X_t}$
1,0	bez transformace

Po transformaci se časová řada rozdělí na krátké úseky, které obsahují 4 až 14 pozorování. V každém úseku se zjistí aritmetický průměr a jejich směrodatná odchylka. Vytvoří se bodový diagram, ve kterém se průměry zanesou na vodorovnou a SE na svislou osu. Pokud body leží přibližně na přímce s nulovou směrnici, nezávisí variabilita řady na úrovni řady a tuto řadu není třeba dále transformovat (blíže [1]).

SARIMA

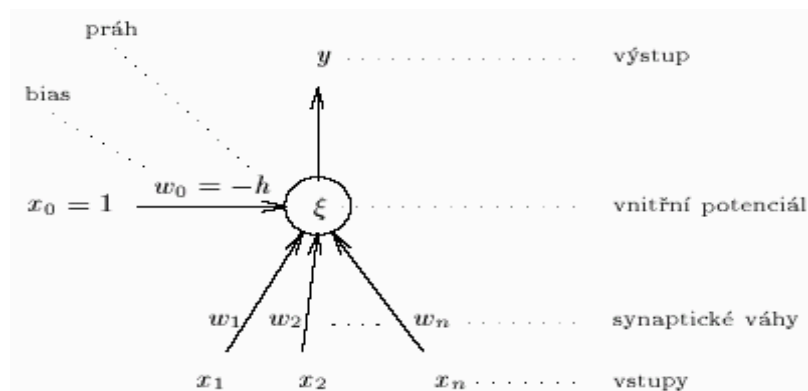
Jak již název napovídá, jedná se o kombinaci AR a MA modelů s diferencováním 'S' značí smíšené sezónní procesy. Přesněji se značí SARIMA(p,d,q)(P,D,Q), kde první závorka značí již zmíněný ARIMA model, druhá závorka obsahuje sezónní ARIMA model. Veškeré vyhodnocování se nemění.

1.1.6 Umělé neuronové sítě

Původním cílem výzkumu neuronových sítí byla snaha pochopit a modelovat, jakým způsobem myslíme a jak funguje lidský mozek. Neurofyzilogické poznatky umožnily vytvořit zjednodušené matematické modely, které se dají využít pro neurovýpočty při řešení praktických úloh z umělé inteligence.

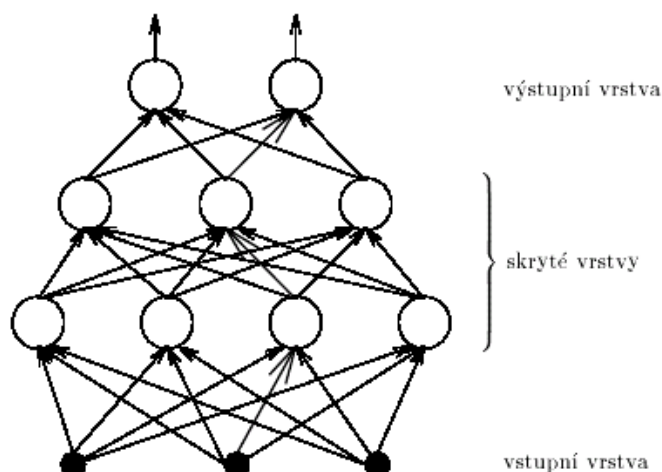
Základem matematického modelu neuronové sítě (dále ANN, Artificial Neural Network) bude formální neuron, který získáme analogicky z neurofyzilogického neuronu, viz Obr. 3.

Vnitřní potenciál $x = \sum_{i=1}^n w_i x_i$.



Obr. 3. Matematický model neuronu [25].

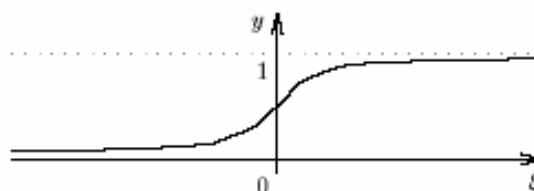
Neuronová síť se skládá z formálních neuronů, které jsou vzájemně propojeny tak, že výstupu neuronu je vstupem obecně více neuronů. Počet neuronů a jejich vzájemné propojení v síti určuje tzv. topologie neuronové sítě. Z hlediska využití rozlišujeme v síti vstupní, skryté a výstupní neurony, viz Obr. 4.



Obr. 4. Vícevrstvá síť [25].

Nejznámější a nejpoužívanější model neuronové sítě je vícevrstvá neuronová síť s učícím algoritmem zpětného šíření chyby (backpropagation), který se používá přibližně v 80 % všech aplikací neuronových sítí. Jiné sítě uvádět nebudu z důvodu rozsahu práce.

Nejpoužívanější přenosovou funkcí, který vyhodnocuje potenciál pro další zpracování je standardní (logistická) sigmoida $y = \frac{1}{1 + e^{-x}}$, viz Obr. 5.



Obr. 5. Standardní sigmoida [25].

Operace předcházející využití konečných výsledků ANN je volba topologie (struktury) sítě, proces učení (aktivní a adaptivní dynamika) a výpočet výstupů pro dané vstupy (aktivní dynamika).

Aktivní dynamika

Po volbě topologie sítě, počítá vícevrstvá síť pro daný vstup funkci $y(w)$, která je určena konfigurací w . V čase $t > 0$ jsou vypočteny vnitřní potenciály

$$x_j = \sum_{i=1}^n w_{ji} y_i, \quad (15)$$

kde

$$y_i = \frac{1}{1 + e^{-lx}}, \quad (16)$$

l je parametr strmosti sigmoidy.

Adaptivní dynamika

Adaptivní dynamika algoritmu backpropagation využívá principu zpětného šíření chyby. Chyba sítě vzhledem k tréninkové množině je definována jako součet parciálních chyb sítě vzhledem k jednotlivým tréninkovým vzorům a závisí na konfiguraci sítě. Vzorec

$$E_k(w) = \frac{1}{2} \sum_{j \in Y} (y_j(w, x_k) - d_{kj})^2 \quad (17)$$

představuje parciální chybu kde y_j je teoretický výstup, d_{kj} reálný výstup, Y počet tréninkových vzorů a

$$\Delta w_{ji}^{(t)} = -e \frac{dE}{dw_{ji}}(w^{(t-1)}) \quad (18)$$

představuje společně s připočtením $w_{ji}^{(t-1)}$ následnou váhu $w_{ji}^{(t)}$, kde $0 < e < 1$ je rychlost učení.

Před učením je vhodně data rozdělit na tréninkovou množinu, validační a testovací. Tréninková představuje zdroje učení, validační je vhodná pro případné poupravení vah, čímž výborně slouží proti overfittingu, k dobré generalizaci, a testovací na ověření správnosti nastavených vah, blíže [9], [10], [14], [15], [25],[26], [30].

1.1.7 Analýza reziduí

Durbin - Watsonova analýza

Uvažujme regresní modely typu

$$y_j = h_j + e_j = b_1 x_{1j} + b_2 x_{2j} + \dots + b_p x_{pj} + e_j, \quad j = 1, 2, \dots, n, \quad (19)$$

kde $x_1, x_1, \dots, x_p$ jsou pevné, nenáhodné, proměnné a $b_1, b_1, \dots, b_p$ jsou neznámé parametry.

Nechť náhodné veličiny e_j lze vyjádřit ve tvaru

$$e_j = Ve_{j-1} + u_j, |V| < 1, \quad (20)$$

kde V je neznámý parametr a u_j jsou nezávislé náhodné veličiny se stejným normálním rozdělením $N(0, s^2)$.

Testujeme hypotézu $H: V = 0$ proti alternativě $\bar{H}: V > 0$. Uvažujme rezidua

$$\hat{e}_j = y_j - \hat{h}_j, \quad j = 1, 2, \dots, n; \quad (21)$$

zde $\hat{h}_j$ je odhad h_j získaný metodou nejmenších čtverců při platnosti H .

Durbin-Watsonova statistika pro test H je definována výrazem

$$D = \frac{\sum_{j=1}^{n-1} (\hat{e}_{j+1} - \hat{e}_j)^2}{\sum_{j=1}^n \hat{e}_j^2}. \quad (22)$$

Podle tabulek, např. viz [16], nalezneme pro n a α dolní a horní meze statistiky D . Je-li $D < D_d$, zamítáme H , je-li $D > D_h$, považujeme nezávislost sousedních hodnot za prokázanou.

Pro hypotézu $H: V = 0$ proti alternativě $\bar{H}: V < 0$ pak stačí D upravit, tedy nové $D^* = 4 - D$. Dále postupujeme stejně jako v předchozím případě, viz [19].

1.1.8 M.S.E. statistika

Zřejmě nejčastějším nástrojem pro volbu mezi několika funkcemi snaží se postihnout reálné hodnoty, je využívána statistika M.S.E. (Mean Square Error). Charakteristika má tvar

$$M.S.E. = \frac{\sum_{t=1}^n (y_t - Y_t)^2}{n}, \quad (23)$$

kde y_t je teoretická hodnota a Y_t je reálná hodnota, viz [12].

1.2 Statistická indukce

Statistická indukce znamená rozšíření závěrů, získaných zpracováním určitého počtu výsledků z výběrového statistického souboru na soubor základní. Jedná se o jeden ze způsobů

zobecnování dosažených závěrů. Při zpracovávání dat z marketingového výzkumu používáme dvě formy statistické indukce:

- odhad,
- statistikou hypotézu.

1.2.1 Odhad

Odhad, který lze provést dvěma způsoby. První spočívá v tom, že z hodnot výběrového souboru vypočítáme jedno číslo (statistiku) a o té potom řekneme, že je odhadem odpovídajícího protějšku v základním statistickém souboru. V tomto případě se jedná o bodový odhad. Příslušnou charakteristiku můžeme odhadnout také pomocí intervalu a poté se jedná o intervalový odhad.

Zdůrazním intervalový odhad průměru:

$$\bar{x} - u_{1-\alpha/2} \frac{s}{\sqrt{n}} < m < \bar{x} + u_{1-\alpha/2} \frac{s}{\sqrt{n}}, \quad (24)$$

kde s představuje výběrový rozptyl.

1.2.2 Statistická hypotéza

Jde o určitý předpoklad o parametrech či tvaru rozdělení zkoumaného znaku. Tento předpoklad se může týkat charakteristik rozdělení náhodné veličiny v základním souboru nebo může být obecnější a vztahovat se pouze k zákonu rozdělení náhodné veličiny (k distribuční funkci, k pravděpodobnostní funkci nebo k hustotě pravděpodobnosti), k náhodnosti, nezávislosti apod.

Předpoklady, které tvoří statistickou hypotézu, se opírají o zkušenosti a dřívější informace a nevycházejí z náhodného výběru. Ten je základem k ověření statistické hypotézy a vlastnímu induktivnímu závěru.

1.2.2.1 Test hypotézy o shodě dvou průměrů

Zde je využita statistika

$$U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \quad (25)$$

kde veličina U má při platnosti H_0 přibližně normované normální rozdělení, s_i^2 představuje výběrový rozptyl, $\bar{x}_i$ je výběrový průměr a n_i rozsah výběru i .

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}, \quad (26)$$

Pokud $U > u_{0,95}$ je na hladině 95% spolehlivosti rozdílnost průměrů tak vysoká, že nelze hodnoty obou skupin aproximovat na společné normální rozdělení.

1.2.2.2 Analýza rozptylu

Často potřebujeme ověřit významnost rozdílu mezi výběrovými průměry většího počtu náhodných výběrů. Toto ověřování významnosti se provádí pomocí analýzy rozptylu. Analýza rozptylu v například v průmyslových aplikacích umožňuje posoudit vliv různých faktorů na výrobní proces, hodnotit vliv použití různých druhů surovin na jakost produkce apod. V marketingovém výzkumu pak můžeme posoudit, zda je odlišná variabilita charakteristik významná v hodnocení respondentů.

Základní myšlenka analýzy rozptylu spočívá v tom, že celkový rozptyl rozložíme na dílčí rozptyly náležející příslušným jednotlivým vlivům, podle nichž jsou empirické údaje roztrženy. Kromě těchto dílčích rozptylů je jednou složenkou celkového rozptylu i reziduální rozptyl, který je způsoben dalšími vlivy, jež ve svém rozboru nepostihuje. Porovnáním složek rozptylu zkoumaného kvantitativního znaku pak určíme vlivy, které významně ovlivňují úroveň tohoto znaku.

Analýza rozptylu při třídění podle jednoho faktoru při stejném rozsahu výběrů

Hypotéza, kterou ověřujeme, se týká středních hodnot základních souborů, na které zvlášť působí určitá úroveň faktoru.

Nulovou hypotézu můžeme zapsat $H_0 : a_1 = a_2 = \dots = a_m = 0$, tj. nulová hypotéza předpokládá, že efekty jednotlivých úrovní sledovaného znaku jsou nulové

Analýza rozptylu při třídění podle dvou faktorů s jedním opakováním

Testujeme hypotézy $H_0^{(1)} : a_1 = a_2 = \dots = a_m = 0$, $H_0^{(2)} : b_1 = b_2 = \dots = b_l = 0$, tj. testujeme tedy hypotézu, že faktory mají nulový vliv.

Analýza rozptylu při třídění podle dvou faktorů s několikerým opakováním

Testujeme $H_0^{(1)} : \sum a = 0$, $H_0^{(2)} : \sum b = 0$, $H_0^{(3)} : \sum (ab) = 0$, kde ab je interakce faktorů.

Tab. 3. ANOVA tabulka pro dva faktory s opakováním.

Rozptyl	Součet čtverců	Stupňů volnosti (v)	Odhad rozptylu
První faktor	$S_1 = n \sum_{i=1}^m (\bar{x}_{i..} - \bar{x})^2$	$m - 1$	$s_1^2 = \frac{S_1}{m-1}$
Druhý faktor	$S_2 = mn \sum_{j=1}^l (\bar{x}_{.j.} - \bar{x})^2$	$l - 1$	$s_2^2 = \frac{S_2}{l-1}$
Interakce	$S_{12} = n \sum_{i=1}^m \sum_{j=1}^l (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})^2$	$(l - 1)(m - 1)$	$s_{12}^2 = \frac{S_{12}}{(l-1)(m-1)}$
Reziduální	$S_r = \sum_{i=1}^m \sum_{j=1}^l \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2$	$lm(n - 1)$	$s_r^2 = \frac{S_r}{m(n-1)}$
Celkem	$S = \sum_{i=1}^m \sum_{j=1}^l \sum_{k=1}^n (x_{ijk} - \bar{x})^2$	$lmn - 1$	X

Tab. 4. ANOVA tabulka pro dva faktory s opakováním, statistika F .

F_i	$F_{0,95}(v_1, v_2)$
$F_1 = \frac{s_1^2}{s_r^2}$	$F_{0,95}(m-1, lm(n-1))$
$F_2 = \frac{s_2^2}{s_r^2}$	$F_{0,95}(l-1, lm(n-1))$
$F_3 = \frac{s_{12}^2}{s_r^2}$	$F_{0,95}((l-1)(m-1), lm(n-1))$

Pokud $F_i > F_{0,95}(v_1, v_2)$ je zamítnuta hypotéza o nulovém vlivu na hladině 95% spolehlivosti, je považován vliv faktoru X za prokázaný.

Studium jsem čerpal hlavně z literatury [4], [5], [6], [8], [11], [16], [17], [18] a [21], které považuji pro popis výroby za hodné dalšího studia.

1.3 Genetické algoritmy

Genetické algoritmy (dále jen GA) představují stochastickou prohledávací metodu, která je inspirována genetickými principy jako je přirozený výběr, křížení a mutace. I když se první pokusy o modelování evoluce na počítači objevily už v padesátých letech, za objevitele genetických algoritmů tak, jak je známe dnes, je považován Holland, který položil základy této disciplíny v sedmdesátých letech.

1.3.1 Princip

Genetické algoritmy se nejvíce používají k řešení problémů optimalizace. Mnoho problémů (například učení neuronové sítě) může být formulováno jako hledání minima či maxima funkce v závislosti na jejích parametrech. V řeči GA většinou mluvím o účelové funkci. Jelikož parametrický prostor je typicky příliš velký pro exhaustivní prohledávání, používáme nějakou stochastickou techniku pro nalezení přibližného řešení. Gradientní metody například reprezentují hledání lokálního minima či maxima pomocí jednoho zpřesňujícího se řešení.

Genetické algoritmy představují jiný přístup, který využívá celou populaci prozatímních řešení, jež paralelně procházejí parametrický prostor a navzájem se ovlivňují a modifikují pomocí genetických operátorů. Tímto mechanismem, nazývaným implicitní paralelismus, se dosahuje jakéhosi synergického efektu, takže populace jedinců najde správné řešení rychleji, než kdyby prohledávali prostor izolovaně. Výsledky ukazují, že populace o n jedincích pracuje zhruba tak rychle jako n^3 izolovaných hledačů. Postup bývá obvykle tento:

1. **Vytvoření počáteční populace**, ze které začíná vlastní prohledávání vytváříme typicky náhodně.
2. **Selekce** – buď může jít o postupnou volbu jedinců nebo o jakkoliv řízenou volbu. Například metodou ‘tournament’, procentním zastoupením nejlepších jedinců (elitářský

mechanismus), metodou ruletová selekce, kde přidává pravděpodobnost na základě úspěšnosti účelové funkce, atd.

3. **Křížení** kombinuje dva vybrané jedince a na jejich základě vytvoří dva či jednoho nového jedince. Nejčastějším typem křížení je křížení jednobodové, kdy se vybere náhodná pozice, na níž se jedinci rozdělí, a jejich části se navzájem vymění. Někdy se používají i vícebodová křížení a jejich limitním případem je ‘uniformní’ křížení, u něhož se pro každou základní jednotku jedince náhodně určí, do kterého z nových jedinců se zařadí. Výhodou uniformního křížení je, že u něj nezáleží na pořadí, v jakém jsou informace v jedinci zakódovány.
4. **Mutace** – zatímco selekce modeluje mechanismus přirozeného výběru a křížení se snaží ze dvou jedinců vytvořit výměnou jejich genetického materiálu jedince kvalitnější, mutace vnáší do populace náhodné změny. Typicky se mutace provádí řádově méně často než křížení. Mutaci realizujeme tak, že náhodně zvolíme pozici v jedinci a danou základní jednotku náhodně změníme. Význam mutace tkví v tom, že dokáže vytvářet zcela nové hodnoty parametrů a zabraňuje tak uvíznutí celého procesu v oblasti lokálního extrému optimalizované funkce.
5. **Výběr** obvykle spočívá v pouhém výběru lepšího ze dvou alternativ. Není problém vytvořit důmyslnější metody, které mohou lépe předcházet nežádoucí a předčasné konvergenci, avšak obecně vyšší univerzalita algoritmu bere svoji daň vyšší náročností na délku trvání hledání optimálního řešení.

Mezi jedny nejpoužívanější a neúspěšnější algoritmy vůbec patří ‘diferenciální evoluce’. Ta se například liší již v selekci, kde nevybírá pro následné operace pouze jedince dva, ale tři, a mutace je velmi úzce spjata s křížením. Přestože algoritmus v práci nepoužiji, zdá se mi vhodné uvést jeden z nejlepších (ne-li nejlepší) univerzálních a velmi jednoduchých algoritmů. Tisíc slov nahradí jedno schéma, viz Příloha P I. Blíže literatura [20], [24] a [28],

1.3.2 Problém obchodního cestujícího

Na mapě je zvoleno N měst a jejich vzájemná vzdálenost pro každou z dvojic. Problémem obchodního cestujícího je určit takové pořadí návštěv jednotlivých měst, aby každé město bylo navštíveno právě jednou, výsledná délka (cena, jízdné, ...) byla nejkratší a cestující se vrátil zpět do města, kde cestu začal.

Náš problém náleží do kategorie NP-úplných problémů, pro které je typická výpočetní náročnost. Např. pro náš případ vždy existuje nějaké ideální řešení (případně více, ale se stejným výsledkem), které můžeme lehce najít tak, že prostě spočítáme délku všech možných cest a vybereme tu nejkratší. Již pro malá N je ale počet vysoký a roste exponenciálně. Časová složitost problému je úměrná počtu všech permutací z n prvkové posloupnosti, tj. je dána výrazem $O(n!)$.

1.3.2.1 Stochastické heuristické metody

V tomto odstavci uvedu stručně rámec tří nepoužívanějších metaheuristik, mírně modifikovaných pro úlohu obchodního cestujícího.

Proto se budeme snažit najít postup jiný, který nám sice se stoprocentní jistotou neumí říci, že zvolená trasa je skutečně nejkratší, ale od ideálního výsledku se výsledek jím předkládaný liší jen málo. Toto tzv. suboptimální řešení je často dostačující.

Genetické algoritmy

GA v problému obchodního cestujícího vyžaduje jiný přístup než obvykle, z důvodu nutného nalezení Hamiltonovy kružnice. Jediný případ, kdy je kružnice okamžitě nalezena je náhodné vytvoření posloupnosti penalizačních hodnot (například kilometry) vázané na přechod mezi dvěma hodnotami (např. dvěma městy). Pak odstranění zmíněné složitosti, vytvoření kružnice, je zachováním posloupnosti křížovatek, kde se nachází právě každá křížovanka pouze jednou. Popíši jen konkrétní změny nutné detaily a odlišnosti k vytvoření algoritmu oproti klasickému použití.

- **Vytvoření počáteční populace** – abychom mohli rychle vizuálně zhodnotit jednotlivé jedince po nalezení optimálního řešení, je vhodné ještě před vytvořením populace určit první křížovanku (město), která ze své pozice po celou dobu řešení neustoupí. Ostatní hodnoty mohou být generovány náhodně, jak tomu je u klasického přístupu.
- **Křížení** také spočívá ve výměně částí jedinců, ale zbytek nově vytvořeného jedince musí být přizpůsoben tak, aby vytvořil kružnici. Pro příklad uvedu dvou-bodové křížení, viz Obr. 6. Konkrétně se jedná se o výměnu dvojice čísel (9,4).

P_1	1	10	7	2	8	9	4	6	5	3
P_2	5	8	2	9	7	4	1	10	3	6
P_1	1	10	7	8	2	9	4	6	5	3
P_2	5	8	2	1	7	9	4	10	3	6

Obr. 6. Dvou-bodové křížení [24].

- **Mutace** má také odlišný charakter. Je sice založena na náhodné změně, ale jako u křížení při povinném zachování kružnice, musí být jedinec doopraven. Literatura [24] nabízí dvě metody mutace, výměnná mutace (exchange mutation) a posuvná mutace (shift mutation). Metoda výměnná mutace nabízí náhodnou výměnu dvou a více genů v chromozomu (jedinci), posuvná změni pořadí jednoho a více genů, viz Obr. 7.

O_1	1	10	7	8	2	9	4	6	5	3
O_1^m	1	10	6	7	8	2	9	4	5	3

Obr. 7. Shift mutation [24].

Simulované žíhání

Simulované ochlazování patří mezi lokální metody. Základní myšlenka je postavena na náhodném prohledávání lokálního okolí daného stavu a následující posun do stavu nového, který má lepší hodnotu optimalizačního kritéria, nebo které vyhoví náhodné funkci připouštějící i zhoršení. Připuštění zhoršení má za následek, že algoritmus nemusí uvíznout v lokálním minimu, jak by tomu mohlo nastat při pouhém zlepšování hodnoty optimalizačního kritéria. Při posunu do nového stavu se prohledávané okolí zmenšuje z důvodu požadavku konvergence algoritmu.

Analogií výše uvedeného algoritmu je představa s atomy. Má-li atom dostatečně velkou energii, může přeskočit energetickou bariéru na hladinu s nižší energií, a se snižující se teplotou se snižuje i pravděpodobnost přeskoků atomu. Atom tak uvízne v místě minimální energie. Pravděpodobnost přeskoků jakožto zhoršení daného stavu je dána vztahem:

$$\text{Exp}(-\delta / T) \quad (27)$$

kde δ vyjadřuje změnu energie a T aktuální teplotu. V případě simulovaného ochlazování to znamená, že při vyšších teplotách budou přijímána větší zhoršení. Při nulové teplotě tudíž nebudou přijímána skoro žádná zhoršení a půjde jen o iterativní vylepšování končící v lokálním minimu.

Metodu simulované žíhání lze zakomponovat nejen do jakéhokoliv běžného genetického algoritmu.

Tabu-search

Metoda tabu-search používá seznam (tabu-list) pevné délky n , ve kterém je uvedeno posledních n změn ohodnocení konkrétně dvojice (proměnná, hodnota). Při každém novém ohodnocení proměnné nějakou hodnotou je tato dvojice zapsána do tabu-listu na začátek a poslední, nejstarší, dvojice je z něj odebrána. Vždy, když se vybírá hodnota pro nějakou proměnnou, zjišťuje se, zda-li není tato dvojice obsažena v tabu listu. Pokud je obsažena, je zakázáno tuto hodnotu proměnné nastavit. Tato technika tedy zabraňuje opětovnému nastavení stejné hodnoty stejné proměnné v nějakém krátkém časovém intervalu, určeném délkou tabu listu. Často bývají zavedena ještě dodatečná (aspirační) kritéria, která nám říkají, kdy může být zákaz ohodnocení dané proměnné danou hodnotou porušen. To může být například tehdy, když nastavením této hodnoty dané proměnné získáme dosud nejlepší nalezené řešení. Jako nejlepší hodnota délky tabu listu bývá často uváděno číslo mezi 10 a 50, i když tato hodnota velmi závisí na struktuře řešeného problému. Metoda rovněž dovoluje umístit svůj princip do genetických algoritmů, blíže [20].

II. ANALYTICKÁ ČÁST

2 ANALÝZA SOUČASNÉHO STAVU A ZÁKLADNÍ ANALÝZA DAT

2.1 Společnost Mitas, a.s.

Akciová společnost MITAS dnešní podoby se skládá ze tří výrobních úseků: v Praze, ve Zlíně a v Otrokovicích.

Tyto závody se zabývají výrobou plášťů pro zemědělské, mimosilniční a stavební stroje, pro motocykly, vysokozdvizné vozíky, pro lehké nákladní automobily a některá letadla. Závod ve Zlíně patří také k významným světovým výrobcům klínových řemenů.

Samotný MITAS Praha byl založen již v roce 1933 v Praze - Strašnicích jako dceřiná firma společnosti Michelin. Michelinu patřily tři čtvrtiny akcií a zbytek vlastnily československé banky. Francouzi přišli do Prahy především kvůli schopné a díky hospodářské krizi dostupné pracovní síle. Nový závod postupně přijal na 250 českých dělníků, kteří zde od roku 1934 vyráběli pneumatiky a duše pro automobily, motocykly a jízdní kola. Rekordní výrobu přinesl rok 1937, kdy se denně vyrábělo asi 300 auto a motoplášťů a téměř 1000 veloplášťů (a duší).

Během války závod přešel pod německou firmu Harburger Phönix, která se orientovala na válečnou výrobu – šlo zejména o pogumování kovových součástí tvrdou pryží. Po spěšném odchodu německého vedení z továrny v květnu 1945 došlo k postupnému návratu k civilní výrobě a posléze ke znárodnění podniku. Na jaře roku 1946 byl název podniku přejmenován na Pneumichelin, n.p. Praha. Na přelomu let 1946 – 1947 byla vyhlášena soutěž na nový název závodu. Zvítězil návrh **MITAS** – spojením Michelin – Veri**TAS**. Firma Veritas vyráběla v Hrádku nad Nisou pryžové obruče na dětské kočárky a v roce 1946 byla začleněna k továrně Michelin. Od léta 1947 se tovární značka MITAS oficiálně užívala.

V letech 1949 – 1952 se začaly instalovat a postupně uvádět do provozu nové stroje a zařízení. Současně probíhaly přípravy na rozšíření výroby středních a těžkých nákladních pneumatik. Výroba veloplášťů přešla pod firmu Rubena Náchod a v roce 1957 v Praze skončila výroba osobních pneumatik. V roce 1958 byl závod přičleněn do nově utvořeného podniku České závody gumárenské, které rovněž sdružovaly společnosti Rubena Náchod a Gumokov Hradec Králové. V roce 1967, kdy skončila rozsáhlá rekonstrukce společnosti, patřil MITAS mezi přední evropské gumárenské společnosti. Postupem doby firma zavedla také výrobu plášťů pro stavební stroje. V březnu 1985 požár zničil nejdůležitější část společnosti

- přípravnou směsí. O dva roky později byla zahájena výstavba nové automatizované míchárny, která byla uvedena do plného provozu v roce 1993. Mezitím se v roce 1990 stal MITAS státním podnikem a posléze akciovou společností.

Tradici výroby pneumatik v dnešním výrobním úseku Zlín společnosti MITAS založil počátkem třicátých let Tomáš Baťa. V lednu 1932 bylo do Zlína dodáno ze zahraničí výrobní zařízení a v dubnu zde vylisovali první pneumatiku o rozměru 32 x 6. O rok později postavil Baťa ve Zlíně pneumatikárnu, která ve druhém pololetí 1934 vyrobila 68 000 pneumatik ve 43 rozměrech. Už v následujícím roce továrna zvýšila výrobu o 100 procent a pláště Baťa se montovaly na všechna tuzemská auta, což výrazně omezilo dovoz zahraničních pneumatik. V roce 1939 Baťa vyráběl 250 tisíc pneumatik ročně a zavedl dále výrobu klínových řemenů, transportních pásů a technické pryže.

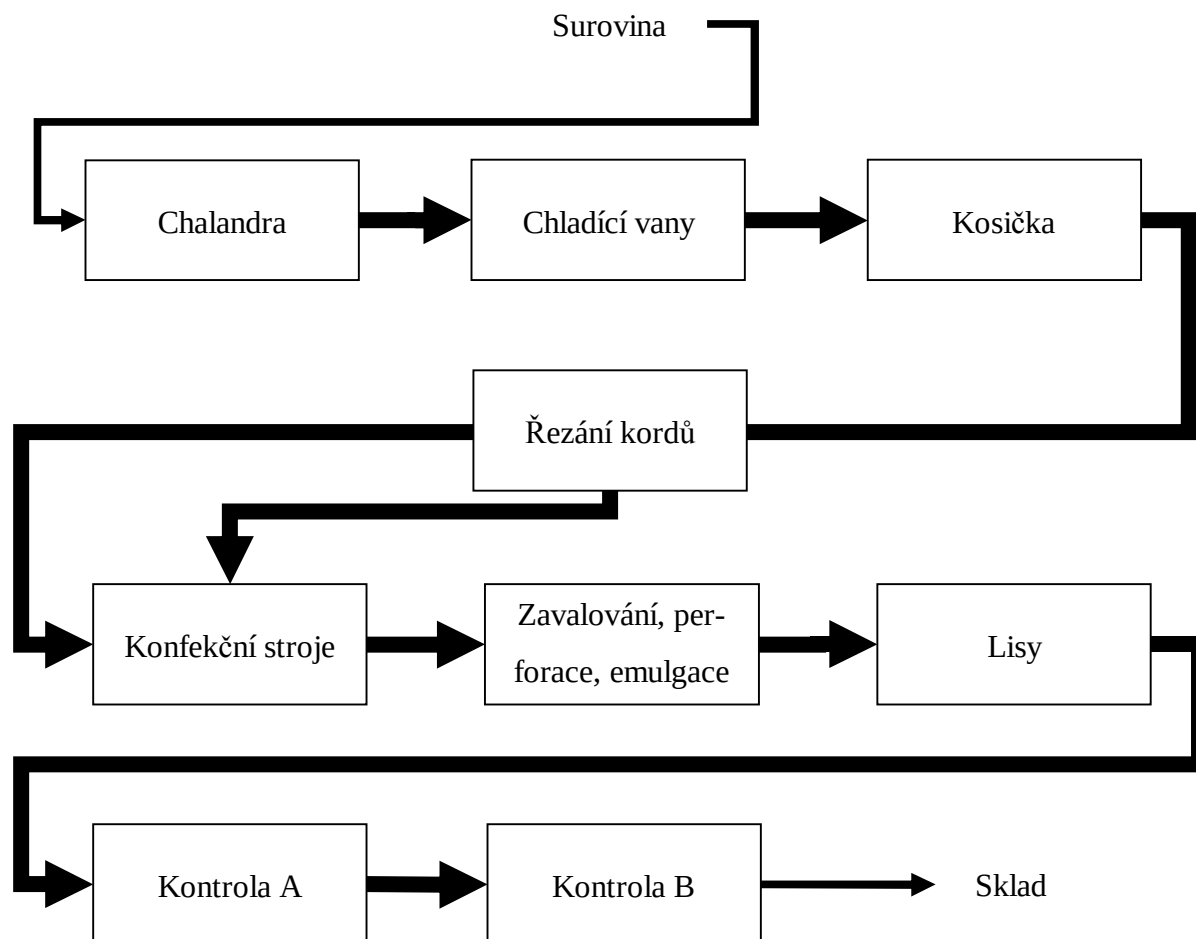
V roce 1945 byla akciová společnost Baťa znárodněna a v roce 1949 se firma Baťa přejmenovává na Svit. Výroba pneumatik se v roce 1953 vyčlenila a vznikl samostatný národní podnik Rudý Říjen. V roce 1967 ve Zlíně vyrobili první osobní radiální pneumatiku a o pět let později byla uvedena do provozu nová hala na výrobu pneumatik v Otrokovcích.

V roce 1991 došlo za účasti společnosti MITAS Praha, Ministerstva průmyslu ČR, Barum Otrokovice, OP Barum Zlín a Motokov Praha k založení akciové společnosti Barum Holding. Většinový podíl v Otrokovické pneumatikárně později získal koncern Continental, zatímco závod ve Zlíně zůstává pod názvem Barumtech (později Beltyr) ve skupině Barum Holding. Z akciové společnosti Barum Holding se postupně formuje Česká gumárenská společnost (CGS) a MITAS Praha se s Beltyrem Zlín slučuje do jednoho subjektu. MITAS již čtvrtým rokem realizuje v obou svých provozech rozsáhlý restrukturalizační a investiční program 'MITAS 2004', na který získal jako jeden z mála českých podniků vládní investiční pobídky. V roce 2002 MITAS zahájil výrobu traktorových radiálních plášťů a MPT radiálních plášťů v celoocelovém provedení [31].

2.2 Výroba motoplášťů

Diplomová práce se zabývá pouze jedné části společnosti Mitas, a.s. Zlín (přiměřený rozsah diplomového projektu to vyžaduje). Objektem snahy se na základě doporučení oddělení průmyslového inženýrství stal právě provoz na výrobu motoplášťů.

Nejlepším způsobem jak popsat jakýkoliv proces je učinit tak nejlépe názorným schématem (a komentáři) znázorňující sled operací, tzv. ‘flowchart’, viz Obr. 8.



Obr. 8. Flowchart provozu na výrobu motoplášťů.

Poznámky k Obr. 8.:

- konfekčních strojů bylo pět, a to konfekční stroj Herbert, Skladačka, RCM, Německý stroj a A60. Každý pro jiný druh motoplášťů,
- tři druhy lisů, celkem 21, přesněji viz Tab. 5.,
- čas obsluhy Chalandy, jako součást Troesteru, závisí na přechodu z jedné receptury suroviny na druhou. Pro každou kombinaci je čas často jiný.

Tab. 5. Přehled lisů v provozu [29].

Počet	Lis – skupina výrobků
5	B-O-M 30" - skútry, bantam, Mich.10"-14"
1	B-O-M 40" - VZV - OFF Taky, motokáry,
15	B-O-M 40" - cestovní, závodní, Michelin

2.3 Vyhodnocení současného stavu aplikace statistických metod

Podobně jako ve své bakalářské práci mohu potvrdit téměř neexistenci aplikace statistických metod, tentokrát v procesu plánování výroby. K vyhodnocování zejména kvantitativních znaků je použita jen nejzákladnější statistika v podobě průměru, váženého průměru, modusu či mediánu. Podle vyjádření jsou aplikovány pouze základní funkce součásti balíků Microsoft Office, Excelu.

V oddělení logistiky dosud nebyla využita ani funkce, často pro pouhé vizuální vyhodnocení, 'přidat spojnicí trendu' v grafu.

V úseku průmyslového inženýrství, pro příklad (ale platí zřejmě obecně), z napozorovaných časových prodlev při výměně receptur mezi dvěma rozměry pracovníci vypočítají průměr, ale dále již nezjišťují interval spolehlivosti, který je velmi, opakuji velmi, důležitý pro následující spolehlivé výpočty v oblasti produktivity práce. Pouhé využití funkce 'popisná statistika' v Excelu může potvrdit naprostou nespolehlivost průměru ze získaných dat. Vypočítaná data prodlev vykazovala průměrnou hodnotu 13,61 minuty; funkce 'průměr' ani nebyla použita, vše vypočteno ručně. Číslo bylo dále zakomponováno do výpočtů, které pracovníci označují jako 'produktivita', ale nikdo nebere v úvahu poznatek ohledně chyby střední hodnoty ve výši téměř 3,4 minuty na hladině 95% významnosti. Chyba je velice blízká hodnotě 25 % z původně získaného průměru, což může v konečném důsledku vést k pomýlení celkem o $\frac{1}{4}$ celkového využitého času! Výsledek a následné odvození na produktivitu pracovníků může později znamenat malé nebo značně velké nároky na zaměstnance.

2.4 Matlab

Vyhodnocení dostupných programů, jejich nabízených statistických nástrojů a způsoby ovládání nejzákladnějších operací, mě dovedlo k závěru, že k vyřešení položeného problému je nejvhodnější software Matlab, který dokáže velmi usnadnit a zefektivnit práci s maticemi.

Nejenže poskytuje kvalitní zpracování matic, ale ve svých rozšířeních, toolboxech, dává k dispozici téměř jakékoliv dostupné algoritmy celého spektra používané především v oblasti matematiky a fyziky. Matlab byl tedy použit na kompletní naprogramování kapitol 2.5.1 až 3.1.4. Více než 500 řádků zdrojového kódu mi snadno umožnilo velmi jednoduché modifikace v závěrečné fázi (při konečné úpravě kódu), kde došlo k různým úpravám z příčiny poopravení dílčích principů (Excel si o takových možnostech může nechat jen zdát). Fáze učení jak pracovat s Matlabem samozřejmě nebyla rychlá.

2.5 Analýza dat základními statistickými nástroji

Původně zdánlivě bezproblémové získání dat z Firmy Mitas, a.s. přinášelo optimismus k splnění cíle alespoň trochu užitečné diplomové práce v praxi. Po snadném získání přibližně 80 % potřebných údajů k vyhodnocení, došlo ke změně stanoviska firmy a od té doby považuje všechna data za citlivá. Nezáskaných 20 % dat představovaly především historické hodnoty objednávek a prodejů motoplášťů po jednotlivých měsících dva roky zpět od září roku 2004. Bylo tedy nutné data vytvořit, aby splňovaly určitá pravidla (například sezónnost, trend, náhodnost). Vycházel jsem z reálných dat potřeby výroby měsíce září 2004 vynásobená součtem určitého koeficientu a náhodného čísla ve stanoveném rozmezí, což zajišťovalo nepodobnost (minimálně o +/- 10 %) a neodvoditelnost reálných dat příslušného měsíce. Přizpůsobení dat v ostatních, přecházejících, obdobích pak bylo zvoleno jako násobky výchozího měsíce s náhodně zvolenými koeficienty v rozmezí 'roz' = +/- 0,33/2 a byla předpokládána roční podobnost pro všech 117 výrobků. Vše splňuje tento zdrojový kód v Matlabu:

```
1 roz=0.33;
2 nahkcoef=1+rand(1,1),25)*roz-roz/2;
3 nahkcoef(:,2:12)=nahkcoef(:,14:24);
4 nahkcoef(:,26)=nahkcoef(:,14);
5 ...
6 for b=2:26
7     realobje(:,b)=realobje(:,b)+b*2;
8 end
```

U každého výrobku pak byla poupravena hodnota o náhodnou změnu v intervalu 24 %, což představuje +/-12 %. Jak již bylo řečeno, potřeboval jsem dvě skupiny dat, prodej a pevné,

známé, objednávky. Prodej jsem pak získal zvýšením objednávek o náhodný koeficient v rozmezí 30 %. Samotné generování dat procházelo častými změnami, aby byla splněna smysluplnost dat, a to zejména pomocí převádění čísel do Excelu a následnému vizuálnímu zhodnocení, proto zde uvádím celý proces, který rozhodně nebyl otázkou deseti minut. Ale nyní již k analýze dat statistickými nástroji.

Vzhledem k tomu, že reálná data by byly beztak poskytnuty maximálně dva roky zpět, využívat složité statistické nástroje (např. umělé neuronové sítě, metody založené na autokorelaci) není nezbytné, avšak pro názornost mohu naznačit postup řešení pomocí umělých neuronových sítí a SARIMA modelování. Pro řešení, předpověď prodeje, jsem zvolil tři metody, a to aditivní model konstantní sezónnosti s lineárním trendem, lineární regresní model a harmonickou analýzu. Samotné použití metod, samozřejmě, předcházely vhodné úpravy dat pro následné zpracování. *V analytické části zmiňuji jen fázi přípravy dat pro další zpracování a aplikuji nejzákladnější statistické nástroje pro analýzu dat.*

2.5.1 Očištění dat o nepravidelnosti

První krok k úpravě dat je očištění o meziroční a meziměsíční nepravidelnosti. V podstatě jde o nepravidelnosti v počtech dní v měsících s uvažováním víkendů a svátků. Nejlépe bude spočítat přesný čas činnosti provozu pro každý den v měsíci, spočítat pracovní čas pro každý měsíc a využít výsledky jako váhové koeficienty. Společnost Mitas, a.s. mi poskytla denní harmonogram, tím mi velmi usnadnila práci, viz Obr. 9.

Su.10.týden	6,67	Su.14.týden	6,67	Su.18.týden	5,67	Su.23.týden	6,67
8 PO	1,00	5 PO	1,00	3 PO	1,00	7 PO	1,00
9 ÚT	1,00	6 ÚT	1,00	4 ÚT	1,00	8 ÚT	1,00
10 ST	1,00	7 ST	1,00	5 ST	1,00	9 ST	1,00
11 ČT	1,00	8 ČT	1,00	6 ČT	1,00	10 ČT	1,00
12 PÁ	1,00	9 PÁ	1,00	7 PÁ	1,00	11 PÁ	1,00
13 SO	1,00	10 SO	1,00	8 SO svátek		12 SO	1,00
14 NE	0,67	11 NE *		9 NE	0,67	13 NE	0,67
Su.11.týden	6,67	Su.15.týden	6,00	Su.19.týden	5,67	Su.24.týden	6,67
15 PO	1,00	19 PO svátek		10 PO	1,00	14 PO	1,00
16 ÚT	1,00	13 ÚT	0,67	11 ÚT	1,00	15 ÚT	1,00
17 ST	1,00	14 ST	1,00	12 ST	1,00	16 ST	1,00
18 ČT	1,00	15 ČT	1,00	13 ČT	1,00	17 ČT	1,00
19 PÁ	1,00	16 PÁ	1,00	14 PÁ	1,00	18 PÁ	1,00
20 SO	1,00	17 SO	1,00	15 SO	1,00	19 SO	1,00

Obr. 9. Výřez z ročního harmonogramu činnosti provozu na výrobu motoplášťů [29].

Hodnoty v třetích sloupcích na konci měsíce sečtu, čímž mi vzniká vytíženost provozu za celý měsíc, jinak řečeno počet dní nepřetržitého provozu v měsíci. Výsledné hodnoty použiji jako váhové koeficienty, kterými podělím prodej za daný měsíc a vynásobím průměrem váhových koeficientů. Tím jsem očistil data o měsíční nepravidelnosti.

2.5.2 Očištění dat o sezónní výkyvy

Abychom byli schopni určit vhodný trend, je nutné data očistit o sezónní výkyvy. Jako vhodnou metodu jsem zvolil prostý centrovaný klouzavý průměr s předpokládaným opakováním výkyvů každých 12 měsíců, protože máme sudý počet pozorování. Pokud bych měl více historických dat, bylo by nasnadě uvažovat i vážený centrovaný klouzavý průměr, ale takto budeme předpokládat pouze lineární závislost sousedních dat. Obě metody si jsou beztak velmi podobné ve výsledku. Pak prostý centrovaný klouzavý průměr má pro 24 hodnot tento tvar: $\bar{y}_t = \frac{y_{t-6} + 2y_{t-5} + y_{t-4} + \dots + 2y_{t+4} + 2y_{t+5} + y_{t+6}}{24}$, což v Matlabu

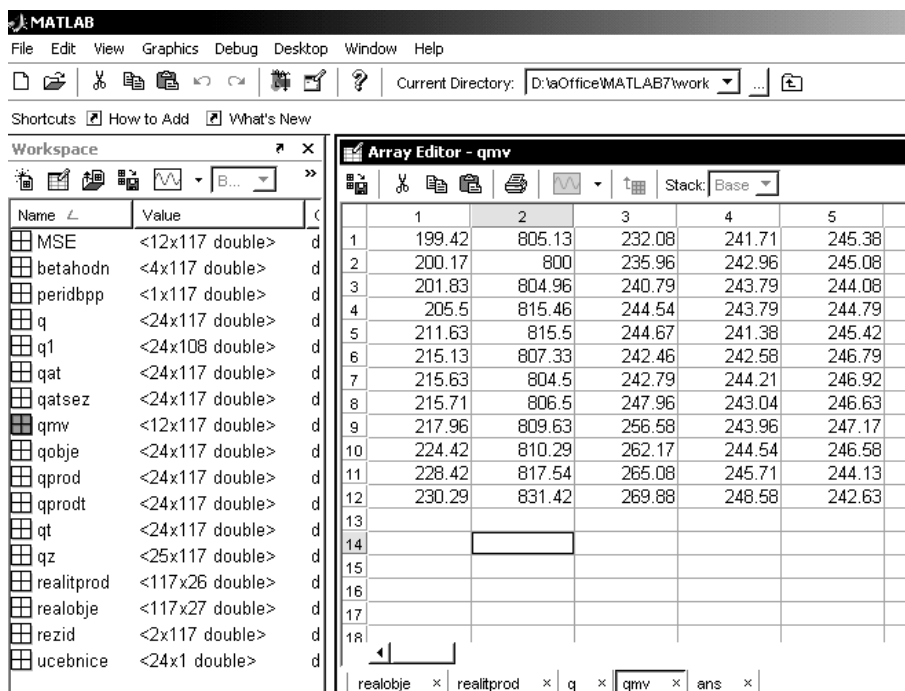
v jednodušší formě znamená:

$$\bar{y}_t = \frac{1}{2} \left(\frac{y_{t-6} + y_{t-5} + \dots + y_{t+4} + y_{t+5}}{12} + \frac{y_{t-5} + y_{t-4} + \dots + y_{t+5} + y_{t+6}}{12} \right), \text{ tedy:}$$

```

1 for b=1:ab(1,2)
2     for a=1:ab(1,1)/2
3         qmv(a,b)=sum(q(a:a+11,b))/24+sum(q(a+1:a+12,b))/24
4     end
5 end, clear a b.
```

‘qmv’ je pak výsledná matice vyhlazených dat, kde ‘a’ představuje období a ‘b’ výrobek, viz Obr. 10.



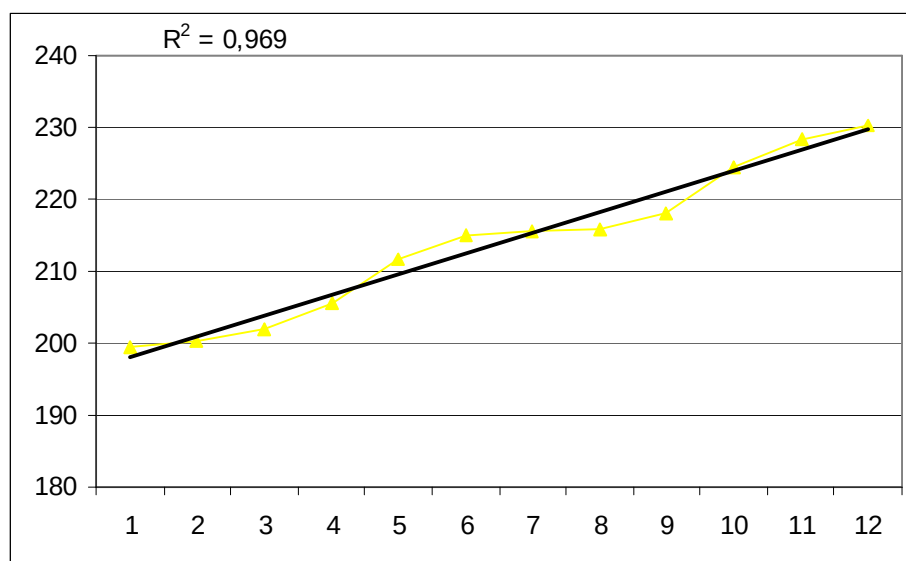
Obr. 10. Matice vyhlazených dat pomocí prostého centrovaného klouzavého průměru.

2.5.3 Nalezení trendu očištěných dat

K nalezení trendu očištěných dat jsem zvolil obyčejný lineární regresní model. Jiné trendy bych nedoporučil již ze zmíněného důvodu malého množství historických dat, což by mohlo způsobit velké výkyvy v předpovědích. Použít můžeme funkci 'LINREGRESE(pole_y;pole_x;b;stat)' v Excelu nebo funkci z Matlabu 'polyfit(x,y,n)'. Ověřovat nalezené koeficienty trendu očištěných dat nepovažuji za vhodné již ze zmíněného důvodu malého množství historických dat, protože by rezidua v kombinaci s ostatními metodami, kde je nutné trend odstranit, při neprokázání vlivu koeficientů značně vzrostly. Navíc se zdá, že u velké většiny je po vizuální kontrole lineární trend vhodný. Pro příklad vyřeším jednu řadu očištěných dat.

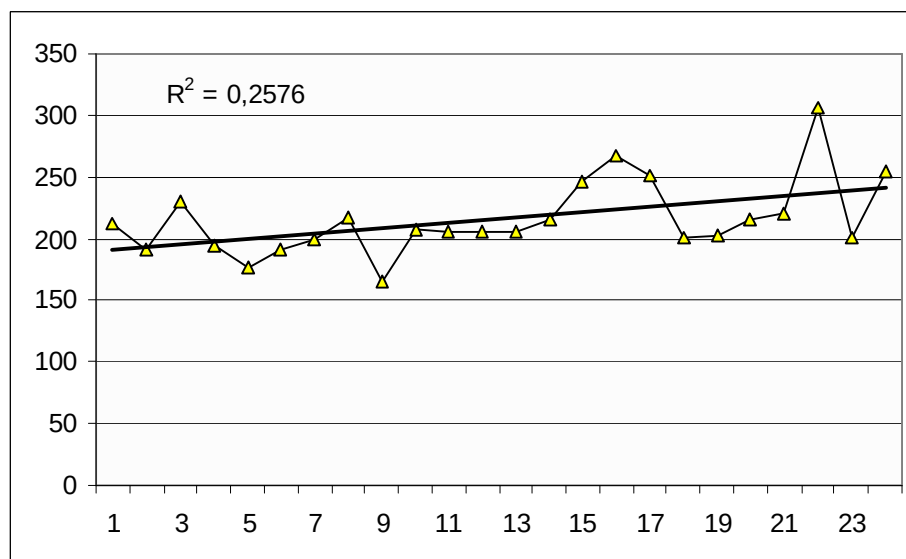
Funkce 'polyfit([1:12],qmv(:,8),1)' pro výrobek pod označením '21212' (první výrobek) našla koeficient $a_1=2,8916$ a $a_0=195,04$, pak rovnice má tento tvar: $\hat{y}_7 = 195,04 + 2,8916 \times (7 - 6)$, kde číslo '7' značí teoretickou hodnotu pro sedmé období a hodnota '(7-6)' je posun, který byl nutný při maticových výpočtech klouzavého průměru v Matlabu, protože očištěná data nejsou pro prvních a posledních 6 období. Pro názornost

v Excelu viz Obr. 11., kde je zřejmá vhodnost daného trendu zjištěná pomocí funkce ‘přidat spojnici trendu’ a zvolením zobrazení R^2 statistiky.



Obr. 11. Matice vyhlazených dat pomocí prostého centrovaného klouzavého průměru.

Z původních neočištěných dat se nám znovu ukazuje síla klouzavých průměrů, kde lineární regrese vysvětluje pouze 26 % hodnot, viz Obr. 12.



Obr. 12. Původní neočištěné hodnoty.

2.5.4 Nalezení trendu prodejů v závislosti na pevných objednávkách nástrojem lineární regrese

Nyní již k prvnímu konečnému řešení určení validní trendové funkce historických dat. Z již mnohokrát opakovanému důvodu jsem zvolil lineární regresní model, kde objednané množství vysvětluje konečný prodej. Jelikož jde o konečné řešení, zařazuji test koeficientů v přecházejícím případě pomíjen.

Funkcí 'regstats(responses,data,model,whichstats)' v Matlabu a následného spočtení t-statistiky pomocí funkce 'tcdf', která pracuje jen v intervalu $(-\infty, x)$, vypočtu koeficienty b a pravděpodobnost omylu hypotézy $H: b_j \neq 0$. Nalezené hodnoty b pak spočteme takto: 'regstats(qprod(:,33),qobje(:,33),'linear',{'beta','covb'})', kde 'qprod' značí závisle proměnnou (konečný prodej) a 'qobje' nezávisle proměnnou (pevné objednávky), 'linear' lineární model, 'covb' kovarianci regresních koeficientů potřebných pro odhad pravděpodobnosti t-statistiky. Pravděpodobnost podle studentova rozdělení vypočteme takto: '2*(tcdf(-abs(regstts{2,1,33}./sqrt(diag(regstts{3,1,33}))), (24-2)))', číslo dvě upravuje statistiku 'tcdf' na interval $(-\infty, +\infty)$, výraz '(24-2)' značí stupně volnosti a výraz uprostřed značí t-statistiku.

Výsledná rovnice má tvar $\hat{y} = 1,199 \times x_{prod} - 4,759$. Vliv koeficientu $b_1 = 1,199$ přijímáme podle t-statistiky na téměř $p = 0,999999999757214$, ale $b_0 = -4,759$ musíme vynechat z důvodu nízké hodnoty $p = 0,2439$, rovnici zkrátíme. I přes nepřijmutí hypotézy o vlivu konstanty a jejím následném vynechání se ukazuje model jako nejúspěšnější ze všech tří uvažovaných, viz Obr. 13., kde první tři řádky porovnávají tři modely (Tab. 6. potvrzuje).

Tab. 6. Reálné objednávky, reálný prodej a rezidua vypočtené pomocí teoretických hodnot.

období	1	2	3	4	5	6	7	8	9	10	11	12
objednávky	96	94	128	112	118	139	133	149	145	134	145	134
prodej	111	107	158	125	123	180	160	179	161	137	173	159
rezidua	-4,1	-5,7	4,5	-9,3	-18,5	13,3	0,5	0,3	-12,9	-23,7	-0,9	-1,7
období	13	14	15	16	17	18	19	20	21	22	23	24
objednávky	139	132	144	131	135	148	152	151	154	154	158	153
prodej	173	160	158	162	160	166	166	173	188	172	184	182
rezidua	6,3	1,7	-14,7	4,9	-1,9	-11,5	-16,3	-8,1	3,3	-12,7	-5,5	-1,5

	30	31	32	33	34	35	36
1	8.0193	52.958	192.13	98.667	223.91	4481.4	1352.
2	5.2917	268.88	347.54	276.08	272.46	13851	1170.
3	3.7083	46.208	135.71	133.92	182.21	3933.2	531.2
4	0	1	1	2	1	1	
5	1	0	0	0	0	0	
6	2	2	2	0	2	2	
7	1.9257	6.7977	11.649	9.9331	13.498	62.716	23.04
8	0.29167	0.20833	0.29167	4.8333	0.29167	0.25	0.2083
9	3.4592	11.39	19.453	21.172	22.495	103.41	38.11
10	2	2	2	2	2	2	
11	0	0	0	0	0	0	
12	56	192	174	174	185	980	36

Obr. 13. M.S.E.

2.5.5 Nalezení trendu prodejů modelem konstantní sezónnosti s lineárním trendem

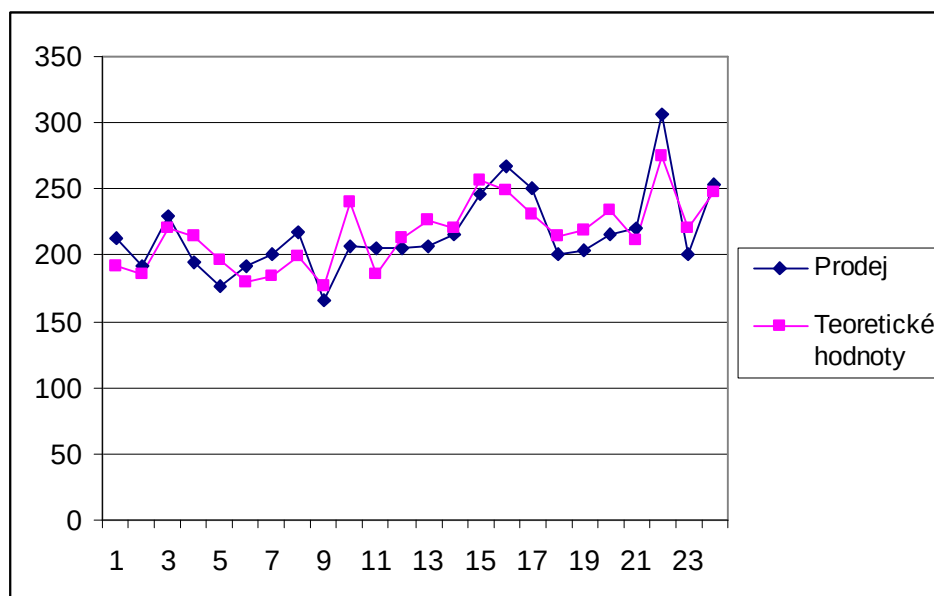
Aditivní model konstantní sezónnosti obsahuje tři koeficienty. Dva z koeficientů jsou již vypočteny v kapitole 2.5.4, třetí koeficient obsahuje sezónní složku, kterou představuje průměr hodnot rozdílu jednotlivých měsíců a trendu očištěných dat. Již je zřejmé zvolení modelu v kombinaci s lineárním trendem.

Pro výrobek pod označením '21212', první výrobek, pro březen 04 máme rovnici:

$$\hat{y}_7 = 195,04 + 2,8916 \times (7-6) + (200-195,04 + 2,8916 \times (7-6) + 215-195,04 + 2,8916 \times (7-6+12)) \times \frac{1}{2},$$

kde výraz v závorce vynásobený $\frac{1}{2}$ je průměr odchylek od trendu očištěných dat, výraz před

závorkou je trend očištěných dat a $\hat{y}_7 = 184$ představuje teoretickou hodnotu vypočtenou pomocí aditivního modelu konstantní sezónnosti. Porovnání všech období daného výrobku viz Obr. 14. - M.S.E=301,83.



Obr. 14. Reálný prodej a teoretické hodnoty výrobku '21212'.

III. PRAKTICKÁ ČÁST

3 APLIKACE POKROČILÝCH STATISTICKÝCH A NESTATISTICKÝCH METOD A ALGORITMŮ

Třetí část obsahuje ne již základní analýzu “poskytnutých” dat, ale pokročilejší a propracovanější principy statistikou definované jako například harmonická analýza, SARIMA modelování a umělé neuronové sítě společně s následnou aplikací genetických algoritmů.

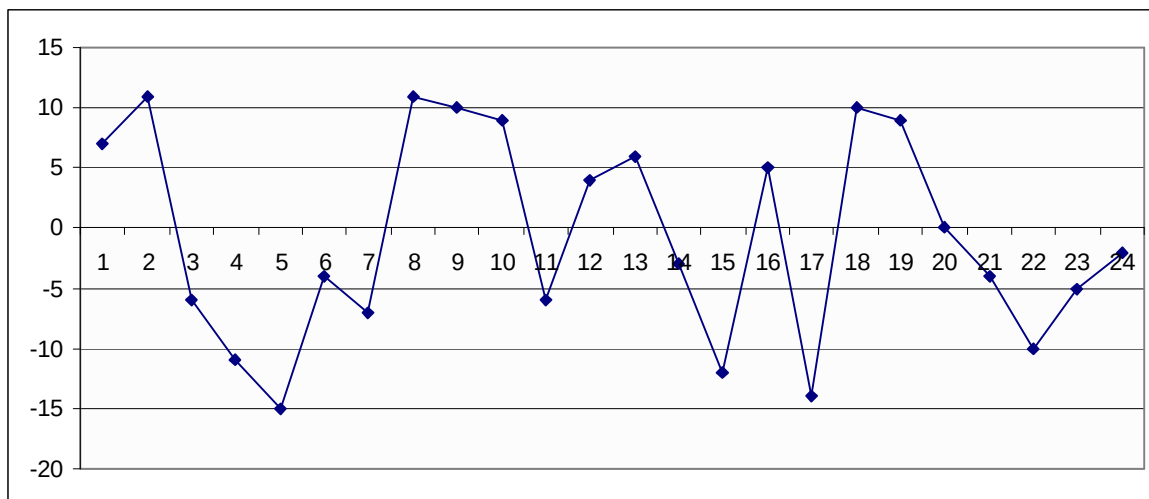
3.1 Časové řady

3.1.1 Aplikace harmonické analýzy

Přestože dat není mnoho, může harmonická analýza odkrýt pravidelné periodické prvky i v řadě o $n = 24$, což se také potvrdilo při vyhodnocování M.S.E. statistiky. V případě lepšího proložení reálných informací datovou řadou oproti trendu, kde nezávisle proměnné jsou reálné objednávky, můžeme slovně vyhodnotit, že trend pro výrobu daného výrobku je spíše v závislosti na trhu než na dodavatelích, kteří se trhu příliš nepodobají. Pokud je úspěšnější statistika aditivního modelu, který mnoho nenaznačuje, pak pořadí úspěšnosti harmonické analýzy a regrese v závislosti na odběratelích může také napovědět, na čem je prodej výrobku pravděpodobněji závislý.

3.1.1.1 Odstranění trendu

Jednoduchá harmonická analýza odhaluje pravidelné periody, které, pokud budou podle nějakého trendu rostoucí, nemohou být spolehlivě měřeny. Samotné periody jsou tedy předpokládány jen jako pravidelné výkyvy bez ohledu na celkový trend. Po odstranění trendu můžeme oddělené periody analyzovat například u výrobku ‘27393’. Trend, odvozený v kapitole 2.5.3, má tvar $\hat{y}_t = 91,292 + 1,489 \times (t - 6)$, pak očištěná data od trendu viz Obr. 15.



Obr. 15. Očištěná historická data výrobku '27393' o trend.

3.1.1.2 Výpočet (odhad) parametrů harmonické analýzy

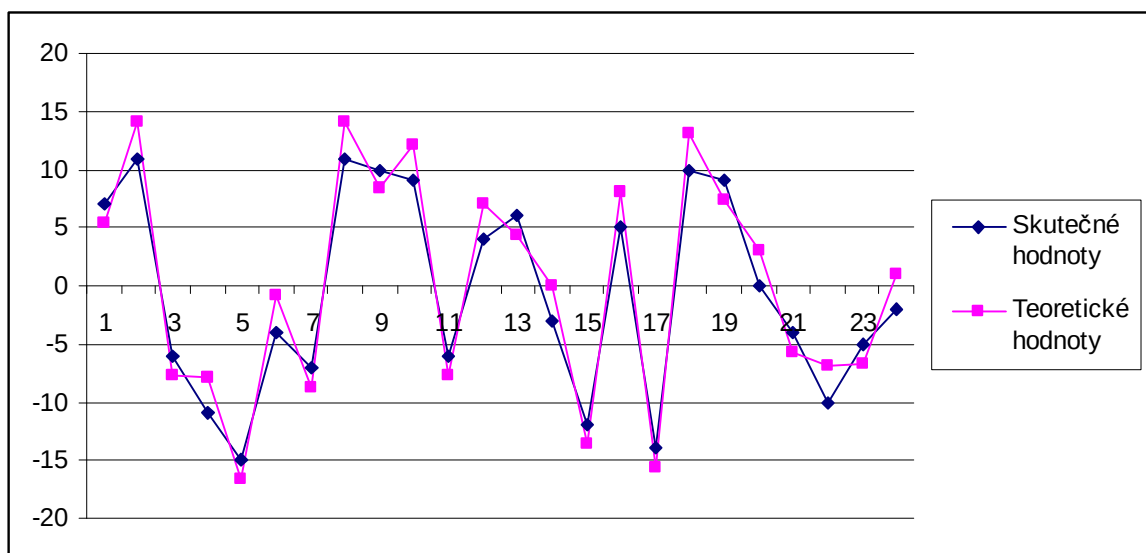
Po očištění dat můžeme pokračovat k vlastní harmonické analýze. Podle vzorců pro výpočet parametrů spočteme jejich hodnoty, a to nejlépe cyklickým způsobem, který nám nadstandardně poskytuje Matlab. Pro osmi měsíční periodu uvádím příklad v Tab. 7., kde jsou vypočteny všechny potřebné koeficienty v posledním řádku.

Tab. 7. Vypočtené teoretické hodnoty pro periodu osmi měsíců .

t	y _t	w	sin(w)	cos(w)	DxB	ExB	Ψ
A	B	C	D	E	F	G	H
1	7	0,7854	0,707	0,707	4,950	4,950	3,980
2	11	0,7854	1,000	0,000	11,000	0,000	4,737
3	-6	0,7854	0,707	-0,707	-4,243	4,243	2,951
4	-11	0,7854	0,000	-1,000	0,000	11,000	-1,360
21	-4	0,7854	-0,707	-0,707	2,828	2,828	-5,671
22	-10	0,7854	-1,000	0,000	10,000	0,000	-7,457
23	-5	0,7854	-0,707	0,707	3,536	-3,536	-5,671
24	-2	0,7854	0,000	1,000	0,000	-2,000	-1,360
Σ	-17	0,7854	0,000	1,000	73,163	17,464	-31,61
x	průměr	x	x	x	a _j	b _j	x
x	-0,71	x	x	x	6,097	1,455	x

Pro všechny periody platí tato rovnice:

$\mathbf{P}_t = -0,71 + 6,1 \times \sin(2 \times p \times 3/24 \times t) + 1,46 \times \cos(2 \times p \times 3/24 \times t) + 5,85 \times \sin(2 \times p \times 4/24 \times t) + 0,13 \times \cos(2 \times p \times 4/24 \times t) + \dots$ Grafické znázornění skutečných a teoretických hodnot při přijmutí všech koeficientů je v Obr. 16.



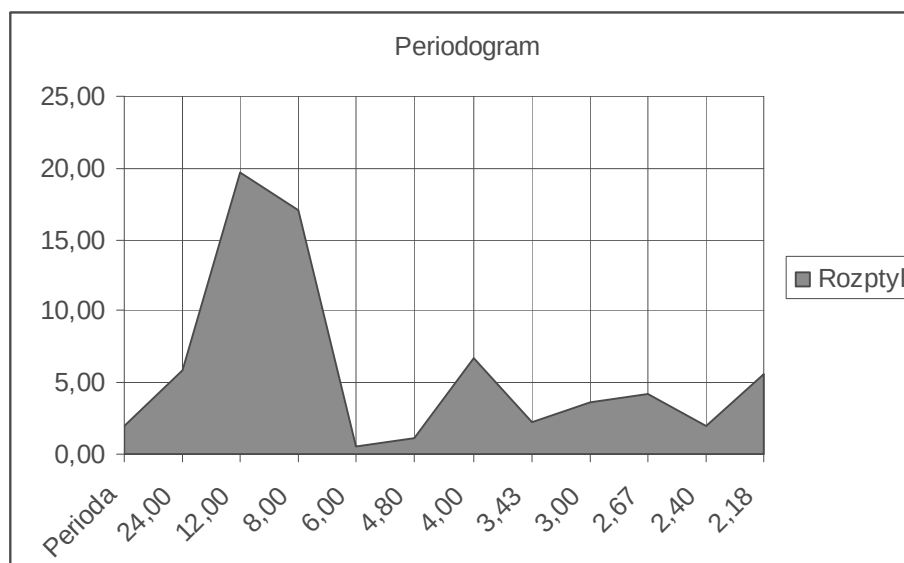
Obr. 16. Proložení skutečných hodnot soustavou sinových a kosinových křivek produktu '27393' očištěných o trend.

Takovýto 'overfitting' samozřejmě nemůžeme přijmout. Rozumnější přístupy ukazují další kapitoly.

3.1.1.3 Periodogram

Výpočet periodogramu, který obsahuje rozptyly teoretických hodnot pro různé periody, nám pomůže vyhodnotit významnost jednotlivých period k celkovému rozptylu napozorovaných hodnot, neboli součtu všech rozptylů pro jednotlivé periody. Hodnota rozptylu pro danou periodu spočteme jako polovinu sumy čtverců koeficientů a a b pro jednukaždou

periodu, tedy: $\text{var}(\mathbf{P}_{t,t}) = \frac{1}{2}(a_{n/t}^2 + b_{n/t}^2)$, kde t je určitá perioda. Pak pro názornost vkládám periodogram produktu pod číslem '27393' v Obr. 17.



Obr. 17. Periodogram výrobu '27393'.

Z obrázku je patrné, že největší rozptyl teoretických hodnot od průměru napozorovaných hodnot je u periody 8, 6, 4, 12 a 2,18, Periody 8 a 6 se na pohled významně vyšší, proto bych se je rozhodl do modelu zařadit. Lepší zhodnocení významnosti poskytuje následující kapitola.

3.1.1.4 Fisherův a Sieglův test významnosti odlišnosti rozptylů v harmonické analýze

Pro statistické posouzení významnosti periodických složek, které nám prozradí významnost s 5% možností omylu, nám pomůže Fisherův a Sieglův test periodicity. Nejprve je vhodné seřadit hodnoty periodogramu od nejvyššího k nejnižšímu.

Fisherův test

Jelikož máme $H \leq 50$, kde $H = n/2 = 12$, pak lze použít aproximaci pro stanovení hladiny selhání pomocí statistiky $P\{W > g_f(0,05)\} \approx 12[1 - 0,05]^{12-1}$. Hodnotu $g_f(a)$ vypočteme jako podíl rozptylu dané periody (postupně od nejvyšší hodnoty) k celkovému rozptylu, následně nejvyšší hodnotu podělíme celkovým rozptylem, ale tentokrát bez již přecházející testované složky rozptylu atd. Tedy $g_{f,j}(0,05) = \frac{var_j}{\sum_{j=1}^m var_j}$ pro první, nejvyšší, hodnotu a

$g_{f,j}(0,05) = \frac{var_j}{\sum_{j=1}^H var_j}$ pro ostatní hodnoty. Konkrétněji viz Tab. 8., kde nenalezneme ani

jednu hodnotu nacházející se pod $P=0,05$. U tohoto výrobku není prokázána existence jediné periody nezávislé na ostatních.

Sieglův test

Vhodnější test pro více významných period je zmíněný Sieglův test, který spočívá v:

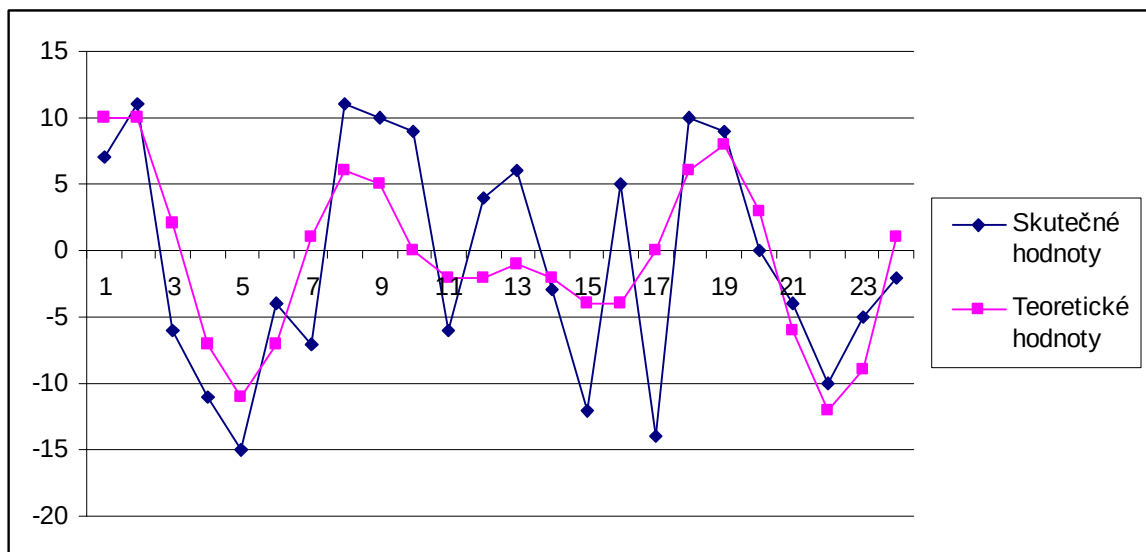
$W_j = \frac{var_j}{\sum_{j=1}^H var_j}$ a statistice $Sw = 0,6 \times g_{f,1}(0,05)$. Pokud je $W_j \geq Sw$, pak budeme považovat

periodicitu za prokázanou. Více v Tab. 8., kde můžeme nalézt prokázanou složenou periodicitu z 8. a 6. periody.

Tab. 8. Fisherův a Sieglův test periodicity.

j	Perioda	Var _j	Fisherův test	Podíl rozptylů	Siegelovo krit.
1	8,00	19,65	0,33	0,278	0,235
2	6,00	17,09	0,19	0,242	0,235
3	3,43	6,69	1,38	0,095	0,235
4	12,00	5,81	1,31	0,082	0,235
5	2,00	5,64	0,94	0,080	0,235
6	2,40	4,21	1,08	0,060	0,235
7	2,67	3,60	0,92	0,051	0,235
8	3,00	2,21	1,35	0,031	0,235
9	24,00	2,00	1,09	0,028	0,235
10	2,18	2,00	0,64	0,028	0,235
11	4,00	1,13	0,68	0,016	0,235
12	4,80	0,58	1,00	0,008	0,235
Σ	x	70,62	x	x	x

Po vyhodnocení můžeme přistoupit ke stanovení hledané funkce. Pro volbu period, pokud byl počet prokázaných period u Sieglůva testu vyšší než dvě, jsem zvolil všechny prokázané hodnoty, a to jak Fisherovým testem, který dokáže velmi spolehlivě určit existenci jediné periody vůči ostatním, tak Siegelovým testem, který je schopen dokázat periodicitu složenou z více period. Pokud byl součet prokázaných period Sieglůvým testem nižší než dvě, pak jsem pochopitelně zvolil periody prokázané pouze Fisherovým testem. Porovnat skutečné a teoretické hodnoty můžeme na Obr. 18.



Obr. 18. Složená periodicitu 8. a 6. periody u výrobku s id '27393'.

Jak můžeme zhlédnout, volba funkce se zdá být vhodná i z důvodu nejnižší M.S.E. statistiky v porovnání s ostatními metodami, viz kapitola 3.1.4.

3.1.2 Kompletní regresní model pro sezónní složku

Tento regresní model také stojí za úvahu. Pracuje na jednoduchém principu, u kterého je jedna závislá veličina a $n-1$ nezávislých, kde n značí počet měsíců v roce. Jako první nezávisle proměnná značí období, ostatní proměnné jsou alternativní, tedy nabývají hodnot 0 a 1 představující jednotlivé měsíce krom jednoho, který bude vysvětlen trendem o jedné proměnné (čas). Ostatní měsíce tvoří funkce o dvou proměnných, trendu a jedné alternativní proměnné. Budeme předpokládat každoroční trendy. Výchozí matice má tvar, viz Tab. 9.

Tab. 9. Výřez výchozí matice pro následné řešení.

y	t	říjen	listopad	prosinec	leden	únor
212	1	0	0	0	0	0
191	2	1	0	0	0	0
230	3	0	1	0	0	0
195	4	0	0	1	0	0
176	5	0	0	0	1	0
192	6	0	0	0	0	1
200	7	0	0	0	0	0
217	8	0	0	0	0	0
166	9	0	0	0	0	0

Pak pro $t = 4$ má funkce tvar $\mathbf{H}_t = b_0 + b_1 \times t + b_3 \times A_3$, kde A_3 nabývá hodnot 0 a 1. Pochopitelně tato metoda je naprosto nevhodná pro periodu jednoho roku, přestože spektrální analýza nezdědka periodu dvanácti měsíců odhalila. Tuto velmi hrubou metodu je v hodné uplatnit s větší datovou řadou, čímž bychom mohli více než dostatečně prokázat jasný opakující se trend alespoň jednoho měsíce.

3.1.3 Durbin -Watsonova analýza náhodné složky

Abychom se ujistili, že náhodná složka opravdu nemá prokazatelnou závislost, alespoň korelační, k tomu nám slouží Durbin-Watsonova statistika, která spočívá v tomto podílu souč-

tu čtverců reziduí: $D = \frac{\sum_{j=1}^{n-1} (\mathfrak{E}_{j+1} - \mathfrak{E}_j)^2}{\sum_{j=1}^n \mathfrak{E}_j^2}$, $D = (0; 4)$ a $H_{p,n}$, $\alpha = 0,05$ pro dolní a horní hra-

nice, kde hodnota pod dolní hranicí znamená zamítnutí hypotézy $H : d = 0, H_1 : d > 0$, kde d je korelační koeficient sousedních reziduí, a hodnota nad horní hranicí přijmutí hypotézy H_0 . Pro testování nepřímé závislosti je $D = 4 - D$ a postupujeme stejně. Uvnitř intervalu není závislost či nezávislost reziduí jasná; pak hranice tolerancí pro různá ' p ' a ' n ' má tyto hodnoty, které dále použijeme:

$$H_{2,24}(d, h) = \langle 1,273; 1,446 \rangle$$

$$H_{7,24}(d, h) = \langle 0,841; 2,03 \rangle$$

$$H_{3,24}(d, h) = \langle 1,188; 1,546 \rangle$$

$$H_{9,24}(d, h) = \langle 0,761; 2,170 \rangle$$

$$H_{5,24}(d, h) = \langle 1,013; 1,775 \rangle$$

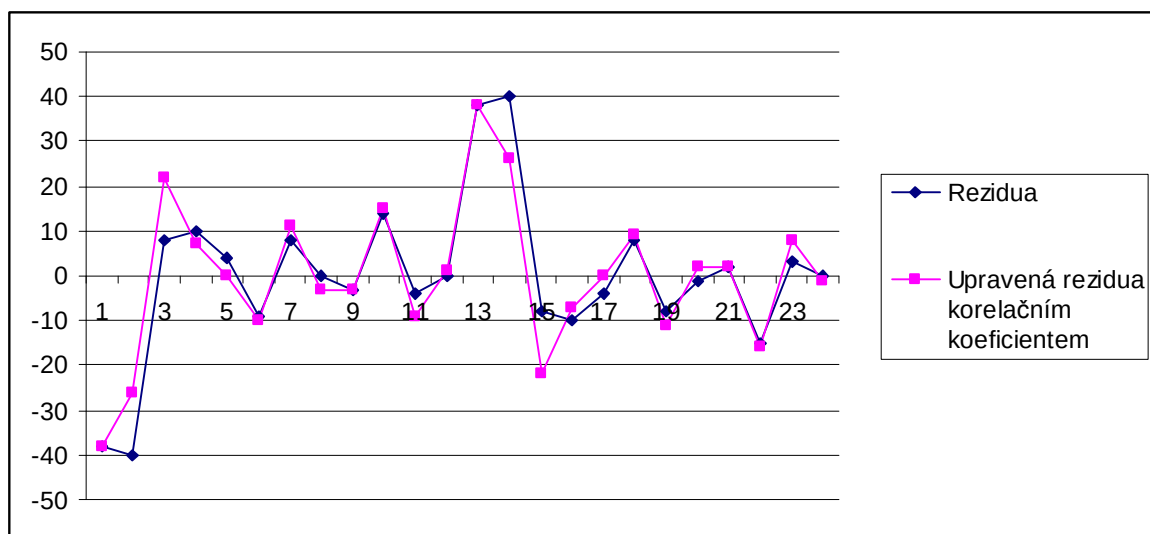
$$H_{11,24}(d, h) = \langle 0,686; 2,223 \rangle$$

Použití pro první dva modely stanovení trendů je jasné - využijeme $H_{2,24}$. Pro příklad uvedu produkt pod označením '26468', u kterého bylo zvoleno řešení pro určení trendu na základě závislosti na prodeji:

$$D = \frac{\sum_{j=1}^{n-1} (\mathfrak{E}_{j+1} - \mathfrak{E}_j)^2}{\sum_{j=1}^n \mathfrak{E}_j^2} = 1,167 < H_{2,24}(d) = 1,273 \text{ pro } d > 0 \text{ a } 4-D = 2,833 > H_{2,24}(d) = 1,273 \text{ pro}$$

předpoklad $d < 0$.

Je zřejmé, je nutno zamítnout hypotézu, že neexistuje přímá závislost sousedních reziduí a naopak považujeme závislost za prokázanou. Protože přijímáme závislost s 5% možností omylu, musíme do teoretických hodnot zakomponovat korelační koeficient $d = 0,356$, který značí vyšší nepřímou závislost sousedních reziduí. Pak konečná rovnice má tvar $\hat{Y}_t = 204,96 + 2,89 \times t + (1,2919 \times t - y_t) \times 0,356$. Porovnání dat před a po využití Durbin-Watsonovy statistiky na rezidua můžeme při pozornějším pohledu shlédnout na Obr. 19.



Obr. 19. Rozdíl v reziduích před a po aplikaci DW statistiky.

Hodnota M.S.E. se snížila z hodnoty 299,21 na 269,82, což není velké snížení, ale po opětovném testu DW statistiky je $D = 1,5994 > H_{2,24}(d) = 1,273$ pro $d > 0$ a $4 - D = 2,4006 > H_{2,24}(d) = 1,273$, což i podle $H_{2,24}(d, h)$ znamená, že je korelace na 95 % odstraněna.

Pro správný cyklus je nutné stanovit podmínky ve zdrojovém kódu:

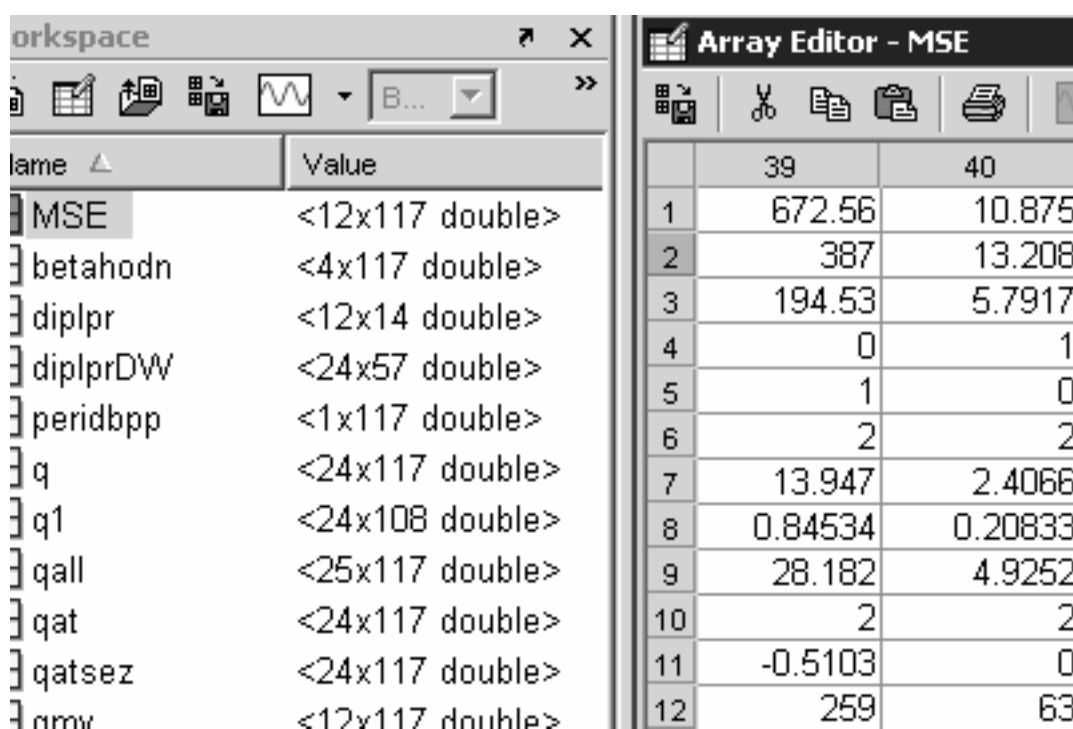
```
1 if DW < 1.273 | 4 - DW < 1.273
2 if DW > 1.446 & 4 - DW > 1.446
```

První řádek obsahuje podmínky pro přijetí hypotézy, kde '|' znamená 'nebo', která je nutná pro zjištění zamítnutí alespoň jedné hypotézy o přímé či nepřímé korelaci. Druhý řádek obsahuje operátor '&', který představuje 'a'. Pokud jsou splněny obě podmínky, pak můžeme s jistotou na 95 % říci, že mezi sousedními hodnotami reziduí není žádná forma závislosti.

3.1.4 Výběr nejvhodnějšího modelu, různé statistiky, predikce a konečná úprava dat

Výběr nejvhodnějšího modelu

Nyní již k poslední fázi, která nám pomůže určit nejvhodnější statistický model pro následnou predikci. Zde není třeba se zahluubovat do statistické literatury a bez dlouhého přemýšlení aplikujeme statistiku M.S.E., neboli chybu střední hodnoty čtverců (mean square error). Vybereme nejnižší hodnotu statistiky M.S.E. u každého výrobku a přiřadíme k němu model. Pro přiblížení mé práce v Matlabu okomentuji Obr. 20.



The screenshot shows the MATLAB workspace on the left and the Array Editor for the 'MSE' variable on the right. The workspace lists several variables: MSE (12x117 double), betahodn (4x117 double), diplpr (12x14 double), diplprDW (24x57 double), peridbpb (1x117 double), q (24x117 double), q1 (24x108 double), qall (25x117 double), qat (24x117 double), qatsez (24x117 double), and nmw (12x117 double). The Array Editor shows the MSE matrix with columns for products 39 and 40, and rows for models 1 through 12.

	39	40
1	672.56	10.875
2	387	13.208
3	194.53	5.7917
4	0	1
5	1	0
6	2	2
7	13.947	2.4066
8	0.84534	0.20833
9	28.182	4.9252
10	2	2
11	-0.5103	0
12	259	63

Obr. 20. Matice výsledků po všech potřebných výpočtech.

Matice MSE(a,b), kde 'b' značí výrobek, obsahuje v prvních třech řádcích statistiky M.S.E., které jsou vyhodnoceny v řádku 4 až 5. Řádky 1 a 4 se vztahují k regresnímu modelu s nezávisle proměnnou představující prodej, řádky 2 a 5 k harmonické analýze a řádky 3 a 6 k regresnímu modelu konstantní sezónnosti s lineárním trendem. Hodnota '2' v řádcích 4 až 6 značí nejnižší statistiku M.S.E., hodnota '1' druhou nejnižší, pokud je u třetího modelu hodnota '2', abychom mohli s jistotou nadsázkou říci, zda je prodej ovládán spíše zakázkami nebo trhem, tedy jestli se trh alespoň trochu liší od trendu minimálně měsíc předem ohlášených nákupů odběrateli. Z obrázku můžeme rozpoznat, že povětšinou by byl nejlépe vhodný regresní model smíšené sezónnosti z naprosto pochopitelného důvodu – malé množství zpětných informací.

Různé statistiky

Řádek '8' obsahuje střední hodnotu reziduí. Je pochopitelné, že se pohybují kolem nuly – tím se můžeme přesvědčit, že výpočty v Matlabu v zásadě nejsou chybné. Malé odchylky od nuly jsou způsobeny zaokrouhlováním na celá čísla.

Řádek '7' ukazuje směrodatnou odchylku chyb a řádek '9' +/- odchylku reziduí, které by měly obsahovat 95 % náhodných složek, podle normálního rozdělení funkcí v Matlabu 'norminv(0,975,průměr,rozptyl)'. Pro výrobek '39', resp. '24649', se 95 % náhodných složek nachází v intervalu (-28,2;+28,2).

Hodnota '2' v řádku '10' ukazuje nezamítnutí hypotézy $H:d=0$, kde $d = \frac{D}{2} - 1$

Durbin-Watsonova testu. Po upravením modelu o autokorelační hodnotu reziduí ($d = \text{řádek '11'}$) se všechny hodnoty nacházejí v prokázaném intervalu určitosti při testování DW statistiky. Jinak řečeno, po splnění povinnosti analyzovat náhodnou složku a odstranění prokázaných závislostí se všechny hodnoty zvolených, nejvhodnějších, modelů rovnají '2'. Durbin-Watsonův test je skutečně dobrým, naprosto nezávislým pomocníkem na vybraném modelu při analyzování náhodné složky.

Predikce nejlepšího zvoleného modelu je v řádku '12'. Výpočet samozřejmě nebyl nepracný a stál minimálně devět řádků plus úpravy zdrojového kódu, ve kterém se člověk znovu zpětně jen velmi nesnadno pohybuje.

Konečná úprava dat

Konečná úprava dat obsahuje přizpůsobení předpovězené výroby harmonogramu na měsíc září 04, a to podělením průměrem koeficientů a vynásobením měsíčního koeficientu, viz kapitola 2.5.1.

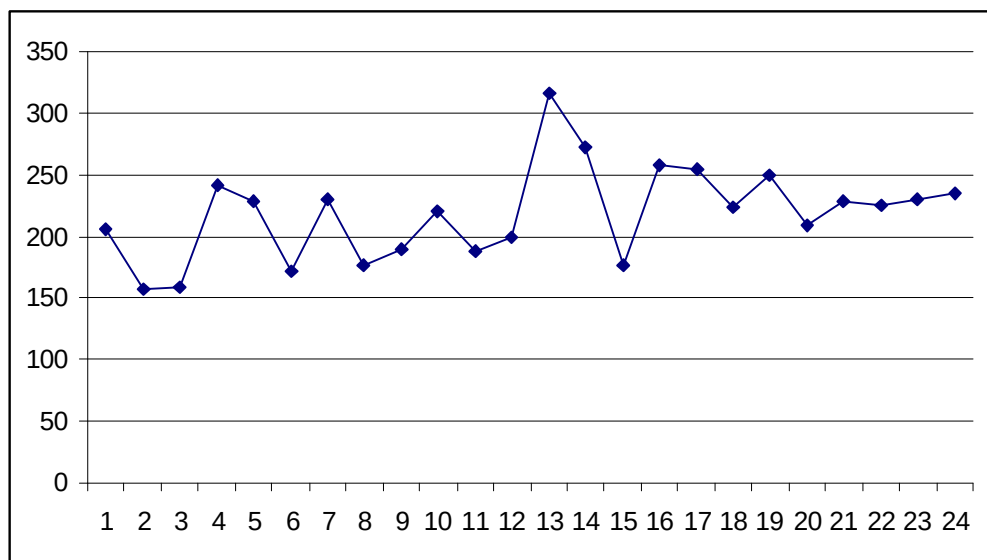
3.1.5 Box-Jenkinsova metodologie

Pro vyhodnocení v případě $N \geq 50$ bychom měli uvažovat i tuto metodologii, která je v praxi, zejména pak v bankovníctví, hojně používána. Jednotný základní postup použití této metodologie je grafická analýza korelogramů autokorelace a parciální autokorelace, případná stacionarizace, zvolení vhodné autokorelační funkce z nabídky AR, MA, ARMA,

SAR, SMA, SARMA, SARIMA, ... a předpověď. Software pro řešení úkolů jsem zvolil anglickou verzi Statistica 6.

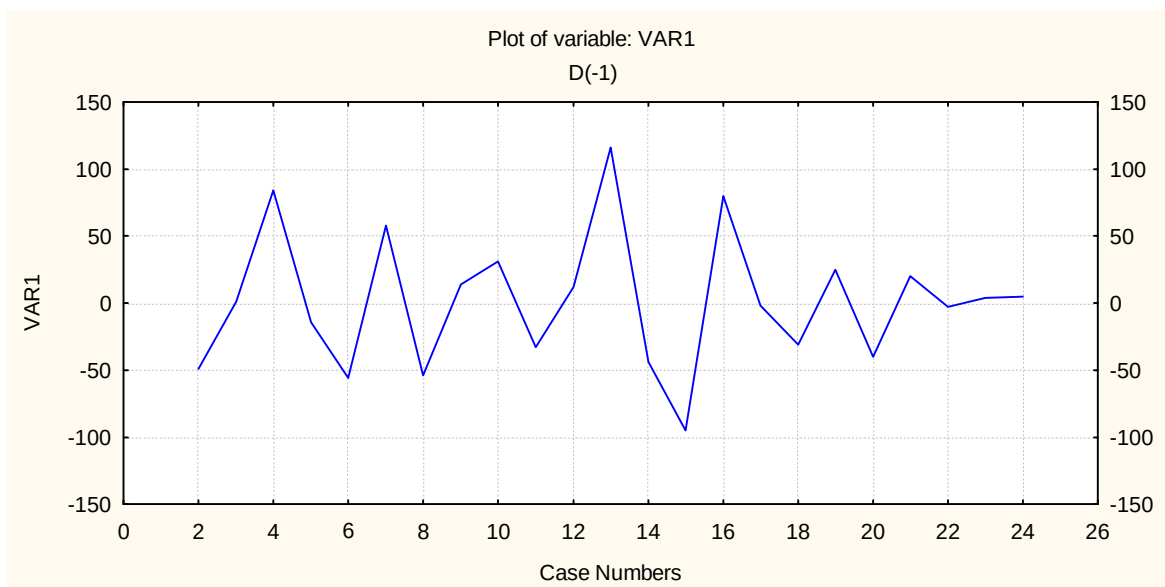
Grafická analýza a stacionarizace

Zkoumat budeme výrobek '26468'. Z grafu na Obr. 21. sice lineární trend není naprosto zřejmý, ale rostoucí trend budeme uvažovat, abychom zůstali u předchozích předpokladů.



Obr. 21. Prodej výrobku '26468'.

Ze závěru z přecházejícího odstavce tímto automaticky volíme model SARIMA(p,1,a)(P,D,Q), ze kterého budeme vycházet. Následně musíme řadu stacionarizovat prvním řádem diferencování. Existují sice přesné metody, například 'rozšířené Dickeyo-vy-Fullerovy testy', ale zabíhali bychom do přílišných složitostí, proto se omezíme na předpoklady. Na obrázku je mírně patrná nutnost stacionarizace v průměru.



Obr. 22. Stacionarizovaná řada hodnot z Obr. 21 řádem 1.

Nyní je potřeba řadu stacionarizovat v rozptylu. Nejdříve přetřansformujeme reálné hodnoty transformací:

$$H(X_t) = X_t^{(l)} = \frac{X_t^l - 1}{l} \text{ pro } l \neq 0, \text{ jinak } H(X_t) = \ln(X_t). \text{ Jelikož se parametr } l \text{ těžko hle-}$$

dá, vystačíme si s porovnáním průměrů a směrodatných odchylek úseků o 4 hodnotách. Pokud budou hodnoty ležet na přímce, kde na ose 'x' je průměr, zvolíme danou transformaci, kde pro:

$$l = -1 \text{ použijí transformaci } 1/X_t,$$

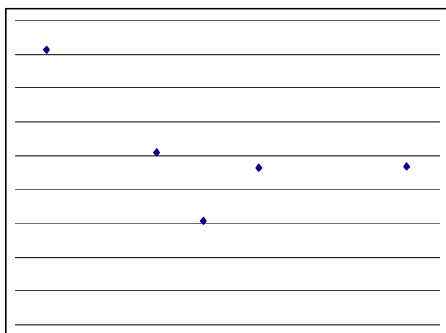
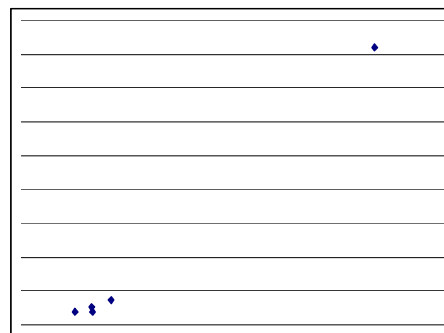
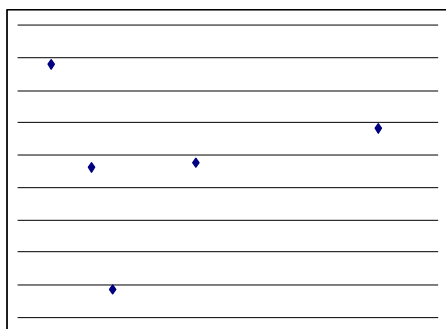
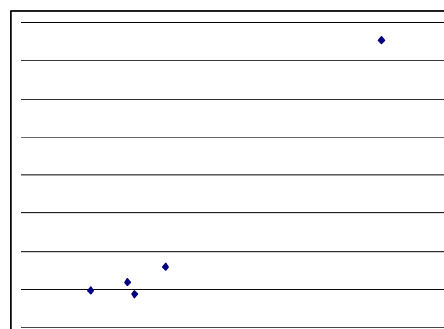
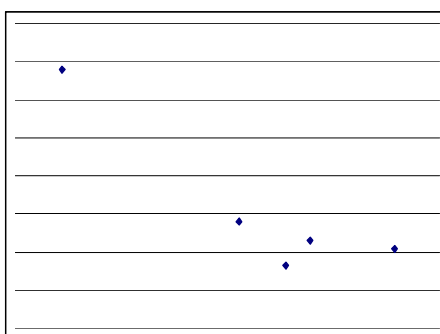
$$l = 0: \ln(X_t),$$

$$l = -0,5: 1/\sqrt{X_t},$$

$$l = 0,5: \sqrt{X_t},$$

a pro $l = 1$ bez transformace.

Řadu rozdělíme na šest úseků, pro každý úsek vypočteme průměr a směrodatnou odchylku. Výsledky jsou v Obr. 23. - Obr. 27.

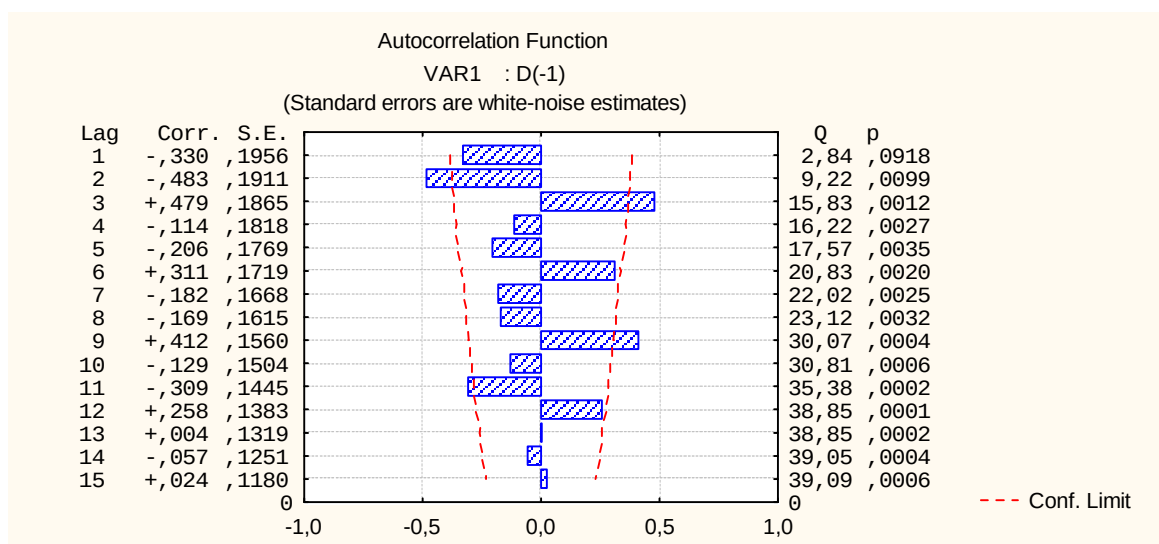
Obr. 23. $l = 0,5$.Obr. 25. $l = -1$.Obr. 24. $l = 1$.Obr. 26. $l = -0,5$.Obr. 27. $l = 0$.

Již z původního obrázku reálných prodejů výrobku '26468' je zřejmé, že není třeba data stacionarizovat v rozptylu, což potvrdil i Obr. 24., kde hodnoty nemají zřetelnou klesající či stoupající tendenci oproti ostatním grafům v obrázcích.

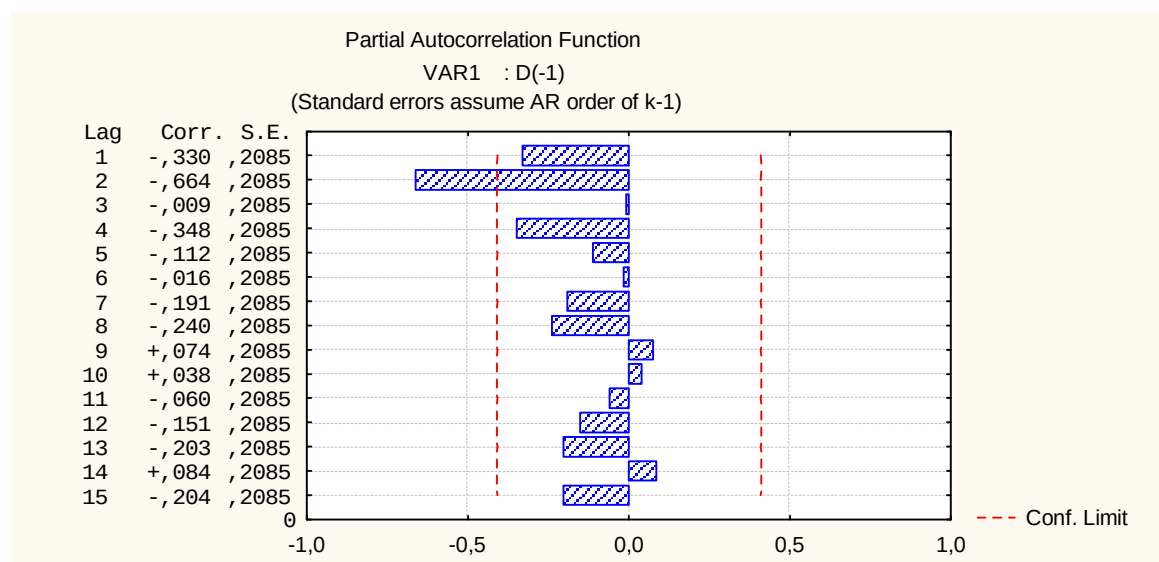
Analýza korelogramů

Výsledek korelační funkce pro jednotlivé kombinace podílů kovariačních funkcí stacionarizované řady můžeme vidět v Obr. 28. Jelikož jde o krátkou časovou řadu, budeme uvažovat více možných autokorelačních funkcí pro následné řešení (i neprokázaných). Obrázek naznačuje vliv dvou předchozích hodnot, tedy AR(2), a opakující se periodu tří měsíců

SAR(1), protože hodnoty vykazují jak v AR(2), tak SAR(1) exponenciální pokles. Obr. 29., ukazující parciální korelační koeficienty (tedy koeficienty korelace bez vlivu vnitřních hodnot mezi kombinacemi jednotlivých hodnot korelačních koeficientů), vykazuje vliv druhého období svým exponenciálním poklesem. Kombinace ACF a PACF v hodnocení periodickou složku tří měsíců SAR(1) nepotvrzuje.



Obr. 28. ACF.

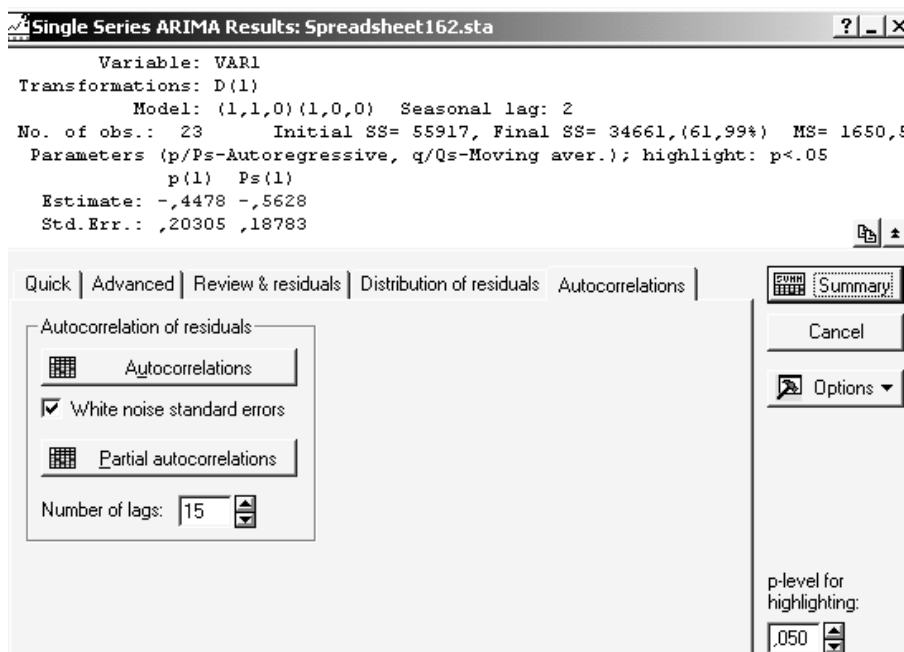


Obr. 29. PACF.

Výběr autokorelační funkce a predikce

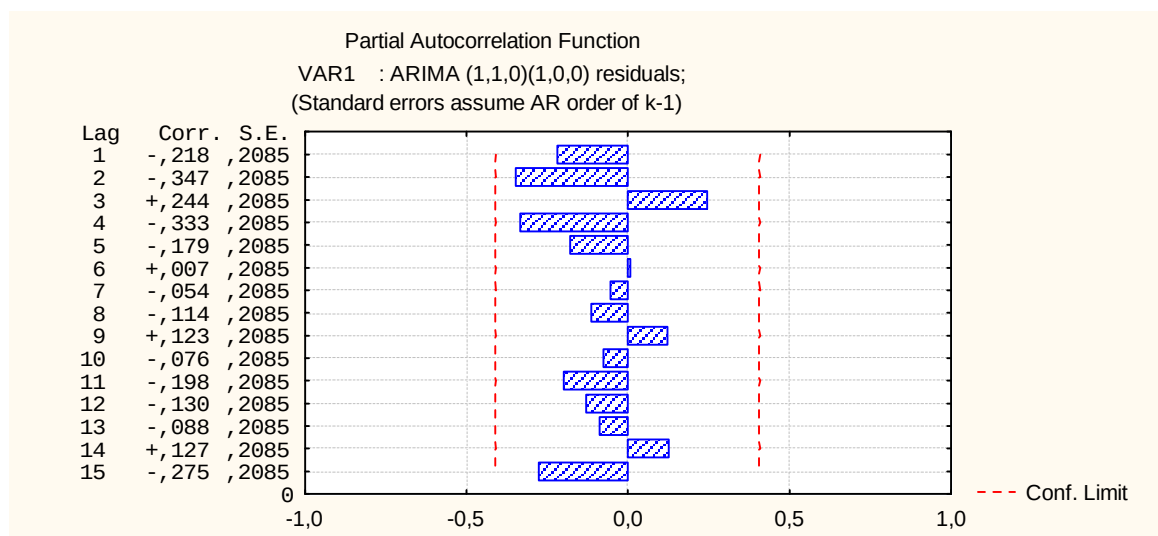
Zvolení správné funkce je otázka především pro zkušeného analytika a nejen z důvodu malého počtu analyzovaných hodnot nepovažuji konkrétně tyto závěry za příliš v praxi použitelné, avšak jako část týkající se mé diplomové práce je podle mého názoru žádoucí.

Analýza korelogramů stacionarizovaných dat mě vedla k vyzkoušení kombinací AR(1), SAR(1) pro opakování dvou, tří a devítiměsíční periody, SMA(1) pro opakování dvouměsíční periody. Nakonec jsem zvolil funkci SARIMA(1,1,0)(1,0,0), tedy ARI(1,1) a SAR(1) pro dvouměsíční periodu. Hledané koeficienty byly softwarových nástrojem Statistika potvrzeny na hladině 95% významnosti, viz Obr. 30. (hodnoty 'Estimate' mají mít červenou barvu jako prokázané).



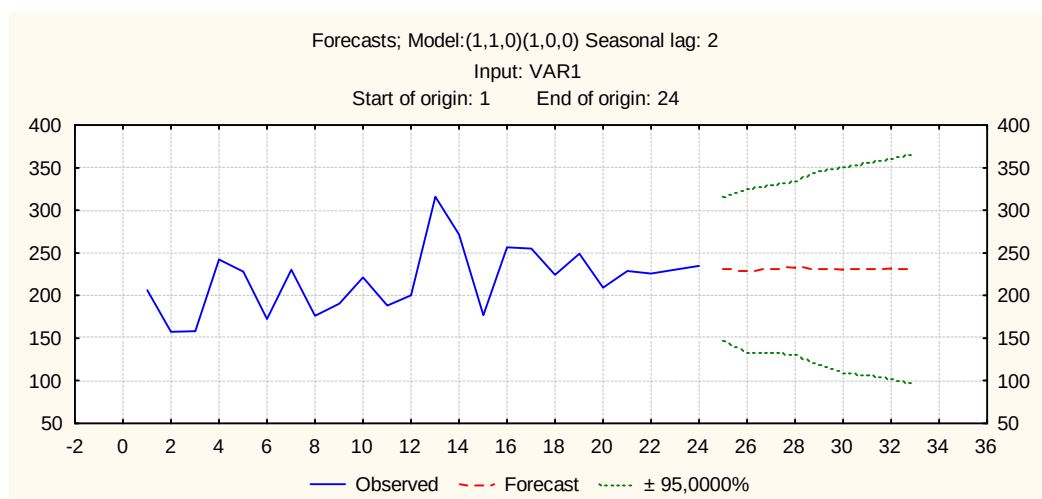
Obr. 30. Výsledek funkce SARIMA (1,1,0)(1,0,0) v SW Statistica 6.

Následně znovu vypočtená statistika PACF potvrzuje mírné odstranění vlivu AR(1) a významné odstranění SAR(1) pro periodu dvou měsíců a vykazuje víceméně náhodné, statisticky nevýrazné hodnoty (viz Obr. 31.)



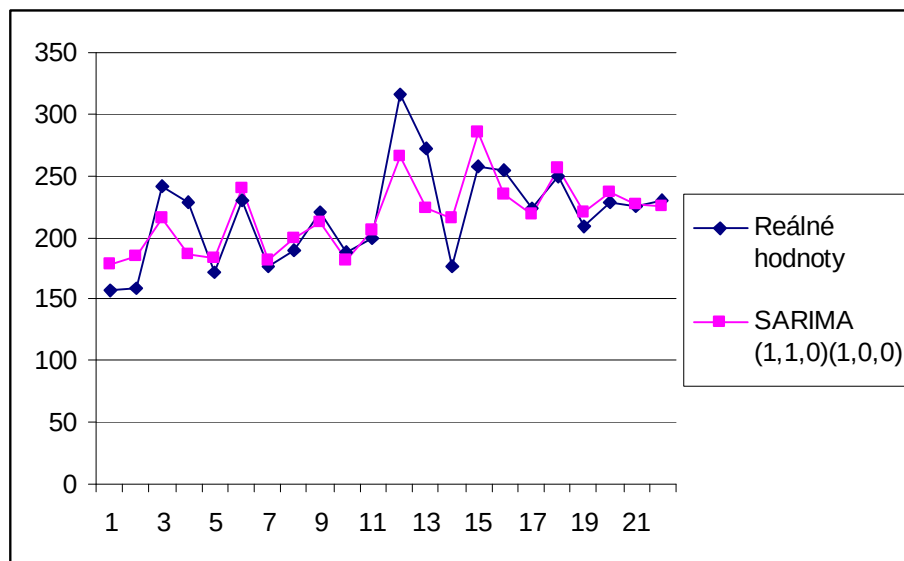
Obr. 31. Kontrola nové tabulky PACF.

Predikce programu Statistica je pochopitelně velmi nespolehlivá, viz Obr. 32.



Obr. 32. Predikce funkce SARIMA (1,1,0)(1,0,0) v SW Statistica 6.

Grafické porovnání reálných a teoretických hodnot můžeme vidět na Obr. 33. Výsledná hodnota M.S.E.=1507,0057 není příliš “mimo“ a můžeme pozorovat velmi dobrou citlivost na následný reálný růst nebo pokles (děkujeme především SAR(1)), ale mnohem lepší statistiku M.S.E. již máme.



Obr. 33. Reálné hodnoty a SARIMA (1,1,0)(1,0,0).

3.1.6 Umělé neuronové sítě a jejich využití pro časové řady

Síla umělých neuronových sítí (dále ANN, Artificial Neural Networks) spočívá ve schopnosti napodobit jakoukoliv spojitou funkci o n proměnných, kde nepřímo rozvoj ANN rozhodl důkazem jen a pouze Andrej Nikolajevič Kolmogorov (*1903). Tato výhoda je hodna užití. Metodu ve svém modelu pro předpověď nepoužiji, protože její potřeby rozhodně nekorespondují s množstvím dat, které máme k dispozici, ale přesto se pokusím pro zvolený produkt nabídnout příkladové řešení (zejména pak z důvodu množství stráveného času nad literaturou s ANN a seznamování se softwarem, podobně jako u harmonické analýzy).

Stanovení předpokladů

Pro správnou konstrukci neuronové sítě musíme vyjít z několika předpokladů, které mohou být pro úspěšnost ANN velmi zásadní, protože tvorba ANN nemá svá přesně daná pravidla, což vyplývá již ze zmíněného důvodu aproximace jakékoliv spojitě funkce (pro nás neznámá), která správně vystihuje datovou řadu. První předpoklad by mohl být, že řada částečně vykazuje závislosti o n hodnot zpět, tzn. má svoji pravidelnou periodu. Druhý předpoklad vyjadřuje možnost částečného odvození od m přecházejících hodnot, třetí by mohl být existence závislosti na určité veličině atd..

Stanovení proměnných

Stanovení proměnných vychází samozřejmě z předcházející kapitoly. První předpoklad, tedy předpoklad period, můžeme přesněji definovat pouhým pohledem na datovou řadu, nebo využít jak služeb spektrální analýzy, tak služeb SARIMA modelování.

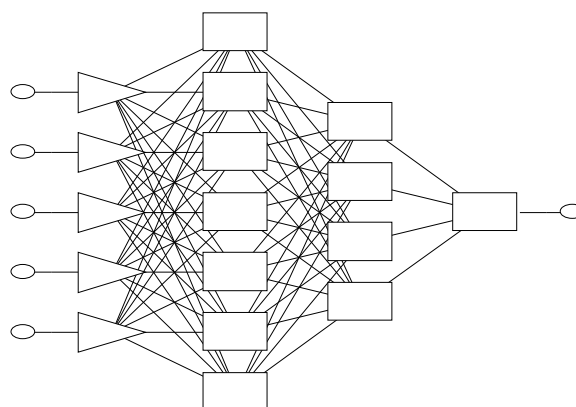
U výrobku pod číslem '23081' jsme v kapitole 3.1.1 prokázali smíšenou 3-měsíční a 2,4-měsíční periodu. Program Statistica po analýze korelogramů naznačil model ARI(2,1), který konečné odhadnuté hodnoty následně prokázal. Model tedy bude vysvětlovat hodnotu v čase X_t hodnotami X_{t-1} , X_{t-2} , X_{t-3} a $X_{t-2,4}$. Hodnota $X_{t-2,4}$ se nachází na 40 % (2/5) "cesty" od hodnoty X_{t-2} k X_{t-3} , tzn., že $X_{t-2,4} = \frac{3 \times X_{t-2} + 2 \times X_{t-3}}{5}$. Pro zohlednění difference

I(1), lineárního růstu, by bylo vhodné zařadit jako další proměnnou klouzavý centrovaný průměr dvanácti hodnot, to však kvůli malému počtu dat není nejvhodnější, nicméně pro kompletnost příkladu jako takového proměnnou uvedu. ANN by si měla váhy nastavit tak, aby významnost nezávisle proměnných byla v souladu s trénovací množinou.

Volba topologie sítě, genetické algoritmy a jiné parametry pro řešení a tvorbu sítě

Tato část je jeden z nejdůležitějších úkolů, podobně jako stanovení proměnných. Sama volba správných proměnných nebo pouze volba vhodné struktury sítě ještě nezaručuje smyslnost výsledků, proto je nutné na tyto dvě operace zaměřit více energie. Jelikož máme 5 nezávisle proměnných, je vhodné zvolit podle literatury o něco více prvků první skryté vrstvy a v druhé vrstvě přibližně průměr první skryté vrstvy a počtu výstupů.

Práci při konstrukci neuronové sítě nám velmi usnadní přítomnost genetických algoritmů zakomponovaných v našem softwaru, Neurosolutions 4.32, který využijeme. Pak tedy síť bude mít minimálně 7 neuronů v první části perceptronu, první skryté vrstvě, a minimálně 4 neurony v druhé skryté vrstvě (vhodný počet neuronů ve skrytých vrstvách v perceptronu určí genetika), poslední část perceptronu bude výstup. Síť bude mít přibližně tento tvar, viz Obr. 34.



Obr. 34. Zvolená topologie neuronové sítě.

Spustím modul ‘Neurosolutions for Excel’, označím sloupce, které jsou vstupy a výstupy, označím jednu množinou dat jako trénovací, jednu jako ‘cross validation’ a množinu dat pro výstup. Dále zvolím ‘train genetic’, kde se objeví další nastavení pro řešení, viz Obr. 35. Počet epoch, vypočtení chyby ANN a znovunastavení vah, zvolím 1500; počet populací 80; počet generací 1000; využiji množinu dat ‘cross validation’, která víceméně zajistí nepřeučení sítě; další nastavení vynechám, protože můj příklad není na klasifikaci dat a není tedy nutné učit druh vzorů, kterých je méně; nechci geneticky optimalizovat (případně vynechání genetikou) vstupy (‘input’), protože jsem již rozhodl o jejich významnosti; kroky učení sítě a momenty učení také nechám na uvážení automatiky; počet neuronů ve skrytých vrstvách zvolím od tří do dvanácti. Další nastavení viz Obr. 36. Volba ‘Tournament’ zajistí nejlepší z nejlepších na základě ‘fitness’, nejlepšího proložení bodů; ‘Heuristic’ zajistí výběr nejlepšího a kombinaci nejlepšího a nejhoršího jedince pro menší pravděpodobnost ustálení na lokálním minimu:

$$Offspring1 = BestParent + r * (BestParent - WorstParent),$$

$$Offspring2 = BestParent;$$

a pro mutaci volíme Gaussovo rozdělení.

Train Genetic

Output Location

Trial Name: Užití neuronových sítí pro predikce

Training Options

Number Of Epochs: 1500

Population Size: 80

Maximum Generations: 1000

Maximum Evolution Time: 60 Minutes

☒ Use Cross Validation

Cross Validation Termination

☐ Terminate after 100 epochs w/o improvement

☐ For Classification problems, make classes evenly weighted

Input Optimization

☐ t-1

☐ t-2

☐ t-2,4

☐ t-3

☒ Průměr

Step Size Optimization

☒ Enable Lower Bound: 0 Upper Bound: 1

Momentum Optimization

☒ Enable Lower Bound: 0 Upper Bound: 1

Processing Element Optimization

☒ Enable Lower Bound: 3 Upper Bound: 12

Help Advanced Genetic Options... OK Cancel

Obr. 35. Train genetic.

Advanced Genetic Options

Progression

☒ Generational ☐ Steady-State

Operators

Selection

Tournament

Selection Basis: ☐ Rank ☒ Fitness

Tournament Size: 2

Crossover

Heuristic

Crossover Probability: 0,9

Number of Tries: 3

Mutation

Gaussian

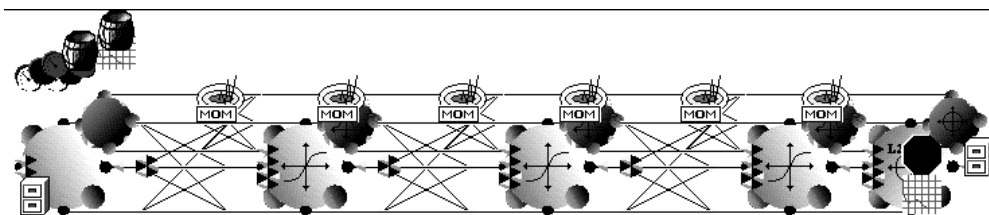
Mutation Probability: 0,01

Help OK Cancel

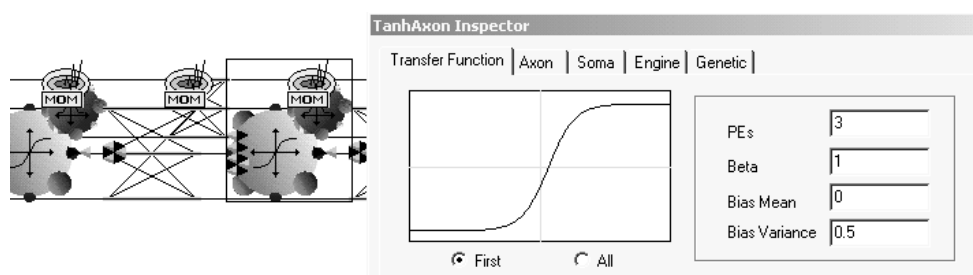
Help Advanced Genetic Options... OK Cancel

Obr. 36. Advanced Genetic Options.

V kompletně připraveném modelu viz Obr. 37. ještě nastavím genetiku pro volbu nejvhodnější strmosti přenosové funkce parametr b , viz Obr. 38.

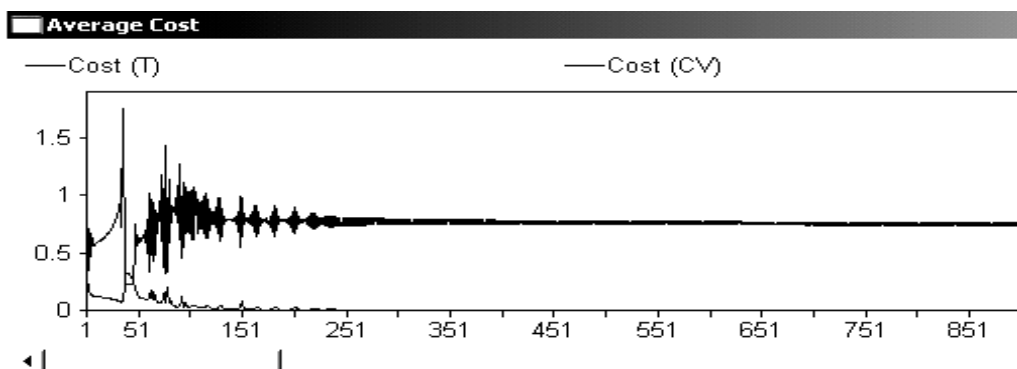


Obr. 37. Model neuronové sítě v Neurosolutions 4.32.

Obr. 38. Parametr b přenosové funkce.

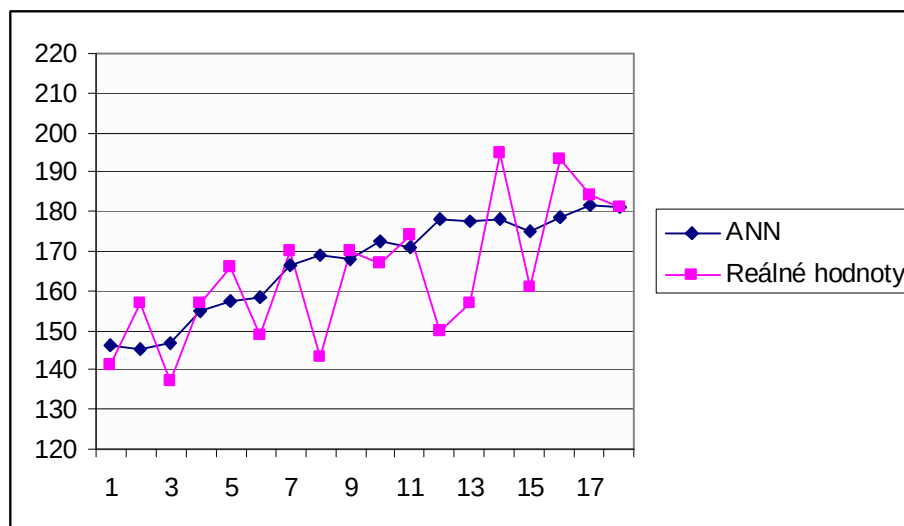
Trénování a výsledky

Pro dosažení věrohodných výsledků testovací množiny, která je v náročnějších případech nutná, je potřeba velmi mnoho času, v ekonomické praxi i několik dní na výkonném počítači (zde se ukazuje zřejmá nevýhoda ANN). V průběhu učení projde síť nespočetným množstvím lokálních minim, přesto nemusí být globální minimum nalezeno. Volba algoritmu učení sítě pochopitelně nebyla složitá. V přibližně 70 % je v praxi používán algoritmus 'Backpropagation' neboli 'strategie zpětného šíření chyby'. V mém případě trvalo učení přibližně 25 minut. Příklad hledání minim chyb pro trénovací a cross validation množiny je na Obr. 39., který ukazuje ne příliš dobrou generalizaci příslušného chromozomu, jedince, svojí oscilací chyb validační množiny. Je také zřejmé velmi rychlé nalezení téměř globálního minima funkce již v 250. epoše u testovací množiny, potvrzuje nízký počet vstupních dat.



Obr. 39 Chyby testovací a validační množiny dat.

Zvolený model tedy pro tréninkovou množinu nevykazuje pochopitelně žádnou chybu, u validační množiny tomu tak není. Konečná hodnota M.S.E. při použití ANN s genetickými algoritmy se pak pohybuje kolem hodnoty 250 oproti o mnoho nižším hodnotám již použitých metod. Testovaný model byl nevhodný. Lépe vystihoval model s následnou menší množinou validačních dat, který alespoň částečně vystihl trend (zgeneralizoval data), viz Obr. 40. a statistika M.S.E. se snížila na 160.



Obr. 40. Proložení dat nástrojem ANN.

Vidíme, že zkušenost pracování s ANN je při vytváření správného modelu zřejmě tou nejdůležitější složkou, např. volba procentuální části validační množiny je pro generalizaci velmi důležitá.

Při nezvolení validační množiny pro křížovou úpravu vah ztrácí nástroj ANN v tomto případě schopnost generalizace, pak výsledky nevykazují naprosto žádné chyby a naznačují ‘overfitting‘.

Závěr

ANN jsou bezpochyby zatím nedostižným nástrojem pro klasifikace dat zvláště v kombinaci s jinými nástroji, například s fuzzy logikou. V našem případě jsme klasifikaci použili v přiřazování významnosti nezávisle proměnných na jedné závislé, jinak řečeno určování vah algoritmem Backpropagation ve vícevrstvé dopředné síti. Nevýhodou je čas strávený nad testováním různých topologií a určování ostatních parametrů pro učení sítě. Universální návod pro sestavení modelu ANN neexistuje, ani nemůže existovat. Statistické zhodnocení

vah a interpretace jednotlivých neuronů je též velmi obtížná. Při přepisu konkrétní nalezené funkce s váhami bychom bez obtíží popsali několik listů a následné zjišťování statistické významnosti by pravděpodobně bylo neúspěšné. Nic na tom však nemění v dalším jistém, bouřlivém rozvoji ve využití pro predikce.

3.2 Statistické nástroje pro popis výroby

Abychom mohli přistoupit k dalšímu kroku, k plánování výroby, je na místě navrhnout statistické nástroje pro popis operací, odhalení závislostí a zajistit hodnověrné informace pro další úpravu statistickými hypotézami. Chci zopakovat, že mým úkolem není vyřešit optimální naplánování výroby, ani popsat výrobu jako takovou. Kapitola je chápána jako nutný dodatek k zachování logické posloupnosti v řešení celého problému, proto zde statistické nástroje budou mít zastoupení v podobě jen těch důležitých a nejlépe dostupných “obyčejnému smrtelníkovi”. Kapitola je silně spjata s kapitolou 2.2.

3.2.1 Volba vhodných statistických nástrojů

Přestože základní statistika není příliš obtížná na pochopení principů, její uplatnění v běžné praxi pracovníka pracujícího s čísly rozhodně běžné není. Proto se omezím jen na základní a jistě velmi dobré pomocníky pro spolehlivá měření.

Odhad spolehlivého průměru

V případě, že chceme odhadnout spolehlivý interval měření, měl by být postup následující:

1. provedeme předběžný výzkum v rozsahu 31 měření například převozu materiálu a spočteme výběrový rozptyl: $s^2 = 1,7$

2. určíme přípustnou chybu od průměru ve výši 0,35, kterou považujeme za spolehlivou a vypočteme nutný rozsah následného měření:
$$n \geq \frac{u_{1-\alpha/2}^2 \times s^2}{\Delta^2} = \frac{1,96^2 \times 1,7^2}{0,35^2} = 90,6.$$

Abychom byli spokojeni s dosaženým konečným rozptylem hodnot, musíme provést celkem minimálně 91 měření.

3. Z nově spočtených výsledků podle $\bar{x} - u_{1-0,05/2} \frac{s}{\sqrt{n}} < m < \bar{x} + u_{1-0,05/2} \frac{s}{\sqrt{n}}$ stanovíme

95% interval spolehlivosti. Pokud chyba nebude splňovat naše očekávání, můžeme přistoupit zpět ke kroku 1, ale již s novými hodnotami.

Test hypotézy o shodě dvou průměrů

Máme dva nezávislé výběry z normálního rozdělení. Chceme zjistit zda rozdíly výběrových průměrů jsou statisticky významné. Nástroj je vhodný například pro porovnání výkonu dvou skupin pracovníků.

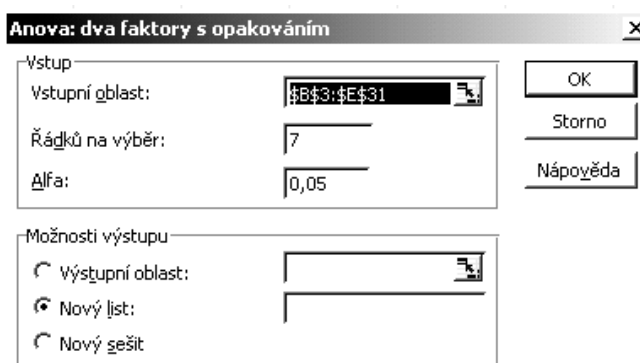
Mějme výsledky měření dvou skupin pracovníků při kontrole motoplášťů. U jedné skupiny byl naměřen průměr 5,1 z 50 hodnot, u druhé pak průměr 4,7, z 53 měření.

$$\text{Pak } U = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5,1 - 4,7}{\sqrt{\frac{1,2}{50} + \frac{1,4}{53}}} = 1,78 > u_{0,95} = 1,64. \text{ Je tedy dokázán významný rozdíl}$$

s 5% možností omylu mezi výkonností skupin.

ANOVA

Pro případ vlivu jednoho nebo dvou faktorů nám skvělé služby nabízí statistický nástroj ANOVA s opakováním. Nástroj je běžně součástí sady Analýza dat v Excelu, proto snadno dostupný, viz Obr. 41.



Obr. 41. Excel – ANOVA s opakováním.

Uvažujme například vliv počtu let zkušeností s prací a vzděláním v počtu chyb při kontrole motoplášťů. Vstupní data viz Tab. 10.

Tab. 10. Tabulka pro nástroj ANOVA.

Praxe (let)	Vzdělání		
	Základní	Vyučen	Maturita
0 - 0,5			
0,5 - 1,5			
2 - 4			
>4			

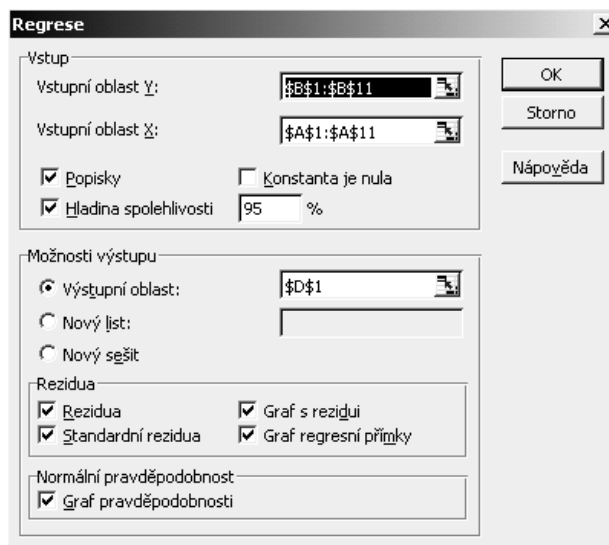
Excel nám poskytne velmi přehlednou tabulku výsledků, z níž můžeme snadno vyčíst, že byl prokázán vliv praxe s možností omylu téměř jedné desetitisíciny chyby, zatímco vliv vzdělání prokázán nebyl, rovněž interakce faktorů nebyla prokázána, viz Tab. 11. sloupec 'P'.

Tab. 11. Výstup nástroje ANOVA v Excelu.

ANOVA						
Zdroj variability	SS	Rozdíl	MS	F	Hodnota P	F krit
Praxe	60,70238	3	20,23413	8,00471	0,000112	2,731807
Vzdělání	3,595238	2	1,797619	0,711146	0,494498	3,123907
Interakce	13,2619	6	2,210317	0,874411	0,51809	2,227404
Dohromady	182	72	2,527778			
Celkem	259,5595	83				

Lineární regrese

Nástroj má skvěle zastoupení v Excelu v nabídce 'Nástroje/Analýza dat'. To je hlavní důvod zařazení do této kapitoly. Pokud tedy máme více nezávislých dat a jednu závislou, můžeme otestovat, zda je vhodné pozorovaná data popsat lineární závislostí, a pokud ano, která. Funkce nabízí vhodné doplňující statistiky k nově vytvořenému teoretickému modelu, která velmi dobře pomohou k rozhodnutí, zda má model takové vlastnosti, aby mohl být popsán lineární regresí, viz Obr. 42. Nejen že nabízí volitelný interval spolehlivosti, ale také poskytuje výstupy s rezidui pro další analýzu nebo různé grafické výstupy pro vizuální zhodnocení. Užitečný může být i graf pravděpodobnosti, který názorně zobrazuje pravděpodobností funkci závisle proměnné.



Obr. 42. Nástroj Regrese v Excelu.

Pro jasnější představu může být lineární regrese aplikována na vhodný stroj, kde mohou mít různé vstupní veličiny vliv právě na výkon. Příklad vstupních veličin může být praxe pracovníka, množství oleje a teplota nastavení stroje, závisle proměnná pak rychlost výstupu.

Tab. 12. Výstup 1.

Regresní statistika	
Násobné R	0,75
Hodnota spolehlivosti R	0,57
Nastavená hodnota spolehlivosti R	0,35
Chyba stř. hodnoty	26,64
Pozorování	10

Z Tab. 12., jako první výstup funkce, vyčteme ne příliš vysokou hodnotu R^2 , pro 95 % a 99% spolehlivost R^2 již naprosto nevyhovuje.

Tab. 13. Výstup 2.

ANOVA					
	Rozdíl	SS	MS	F	Významnost F
Regrese	3	5632,58	1877,53	2,64	0,14
Rezidua	6	4259,52	709,92		
Celkem	9	9892,10			

Tab. 13., hodnotící rozptyly, ukazuje na nepříliš významný rozdíl rozptylů teoretických hodnot a rozptylu reziduí. Významnost hodnoty F dosahuje 14 %, my však jako dostateč-

nou hodnotu považujeme maximálně 5% významnost statistiky F. Teoretický model lineární regrese proto není dobrou volbou.

Tab. 14. Výstup pro hodnocení koeficientů.

	<i>Koeficienty</i>	<i>Chyba stř. hodnoty</i>	<i>t stat</i>	<i>Hodnota P</i>	<i>Dolní 95 %</i>	<i>Horní 95 %</i>	<i>Dolní 99,0 %</i>	<i>Horní 99,0 %</i>
Hranice	487,62	44,28	11,01	0,00	379,26	595,97	323,45	651,79
Úroveň oleje	2,12	5,89	0,36	0,73	-12,29	16,52	-19,71	23,95
Úhel	-2,46	2,53	-0,97	0,37	-8,65	3,74	-11,85	6,93
Praxe	3,12	1,24	2,52	0,05	0,09	6,14	-1,47	7,70

Poslední nejdůležitější výstup z analýzy dat je v Tab. 14., ze které lze vyčíst prokázání vlivu na výkon s 95% spolehlivostí pouze u jedné nezávisle proměnné, a to praxe. Pouze u této proměnné nedosáhla dolní mez pro 95 % spolehlivost hodnoty nula a méně. U spolehlivosti 99 % má na model vliv pouze konstanta a_0 , jejíž přesah směrodatné odchylky (sloupec 'Chyba stř. hodnoty') je více než jedenáctinásobný a je spolehlivě prokázán.

3.3 Aplikace genetických algoritmů v procesů plánování výroby

Pro aplikaci genetický algoritmů jsem se rozhodl z důvodu složitosti výrobního procesu výroby motoplášťů, který dokáže optimalizovat snad jen řízená metoda pokus-omyl. Genetické algoritmy právě tuto metodu více než vylepšily a jsou snad jediným uvažovaným nástrojem pro řešení velmi složitých problémů s velkým množstvím proměnných.

V pozdním čase, opakuji, jsem se dozvěděl o citlivé povaze získaných dat, čímž se můj model nechtěně zjednodušil. Pokusil jsem se o model plánování především na základě úzkého místa, které stanovuje maximální možnou kapacitu výroby. Přestože jde o velmi složitý problém v mnoha závislými, alespoň dílčí řešení pomocí genetických algoritmů není jistě k nepotřebě.

3.3.1 Rozplánování výroby v úzkém místě provozu

K řešení tohoto problému mi velmi pomohl již vyvinutý software Evolver, který je určen pro Microsoft Excel. Excel mi rovněž velmi pomohl, a to svoji podstatou, ohromným zjednodušením při tvorbě závislých dat, která se v problému hojně vyskytují.

Příprava dat v Excelu

Podle reálných dat bylo nutné vytvořit seznam výrobků, přiřadit časy lisování a typ lisu, konkrétně viz Obr. 43.

	A	C	D	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	1	počet lisů BOM 40"							2						
2	5	počet lisů BOM 30"							1	0				0	
3	15	počet lisů BOM 40"							5	0				0	
4									15	0	15	5	1	0	
5															
6	č. výrobku	doba lisování a manipulace	počet lisů	sklad	GA výr.	GA ano/ne	nutná výroba	vyrobeno	potřeba	chybí	počet hodin na lis	počet hodin na lis	počet hodin na lis	celkem hodin	sklad
11	22041	15,3	5	0			1	0	320	320	0	0	0	0	0
12	22043	22,3	5				0	0		0	0	0	0	0	
13	22049	22,3	5				0	0		0	0	0	0	0	
14	22051	15,3	5				1	0	300	300	0	0	0	0	
40	24432	32,3	15				0	0		0	0	0	0	0	
41	24532	15,3	15				1	0	400	400	0	0	0	0	
42	24544	15,3	15				0	0		0	0	0	0	0	
44	24643	22,3	15				1	0	500	500	0	0	0	0	
45	24649	32,3	15				0	0		0	0	0	0	0	
46	24796	22,3	15				1	0	130	130	0	0	0	0	
79	25141	17,3	1				0	0		0	0	0	0	0	
80	25250	17,3	1				1	0	155	155	0	0	0	0	
91	24087	22,3	15				0	0		0	0	0	0	0	
92	24134	32,3	15				0	0		0	0	0	0	0	
93	24167	22,3	15				1	0	300	300	0	0	0	0	
94	24403	32,3	15				1	0	300	300	0	0	0	0	
115	52008	15,3	5				0	0		0	0	0	0	0	
116	52010	15,3	5				1	0	150	150	0	0	0	0	
117	61011	15,3	5				0	0		0	0	0	0	0	
121	52020	17,3	1				0	0		0	0	0	0	0	
124						0	9	0		2,6E+203					
125								0							
126										-2,6E+203					

Obr. 43. Výřez z Excelu.

Přiřazení součtu časů lisování a manipulace a druh lisu, který jsem vyjádřil počtem dostupných druhů lisů (viz sloupec 'D' v Obr. 43.), k danému výrobku jsem provedl pomocí cyklů v Matlabu. Princip řešení spočíval v určení maximálního počtu vyrobených kusů při využití časového fondu 22,5 hodin, tedy tří směn po 7,5 hodinách. Jelikož jde právě o řešení na základě úzkého místa, lisů, které bylo skutečně využito na téměř 100 %, neuvažuji další veličiny (neuvažuji také z důvodu nedostatku "citlivých" dat). Dále princip požaduje, aby bylo možné stanovit výrobu produktů, které je nutno bezpodmínečně vyrobit. To splňuje sloupec 'U' ('potřeba'). Programu Evolver byl nabídnut sloupec 'Q' pro stanovení počtu vyrobených kusů zvoleného produktu. Sloupec 'R' byl také nabídnut Evolveru, aby binárně rozhodl o skutečné výrobě položky výroby z důvodu lepšího a rychlejšího nalezení řešení okamžitým zvolením takového výrobku, který vyplní kapacity volného lisu bez uvažování Celých čísel ve sloupci 'Q', což by mohlo vytvářet závislosti zpomalující účinnost algoritmu.

Protože sloupec ‘U’, tedy sloupec, který je uživatelem ručně nastaven a určuje nutnost výroby daného výrobku (je zřejmě ten nejdůležitější), bylo potřebné přizpůsobit model, aby zohlednil tuto skutečnost a popřípadě zvýšil vhodnou váhou její důležitost. Řešením byl sloupec ‘S’ (‘nutná výroba’), od které se odvíjí již konečná reálná budoucí výroba, sloupec ‘T’. Pak je nutné zohlednit jak Evolverem určující výrobu položky (sloupec ‘R’), tak nezbytnou potřebu vyrobit takové množství, aby byl minimálně splněn sloupec ‘U’. Konkrétně vysvětlený princip ve sloupci ‘S’ obsahuje tento příkaz:

$$=KDYŽ(R11=1;1;KDYŽ(V11>0;1;KDYŽ(U11-P11>0;1;0)))$$

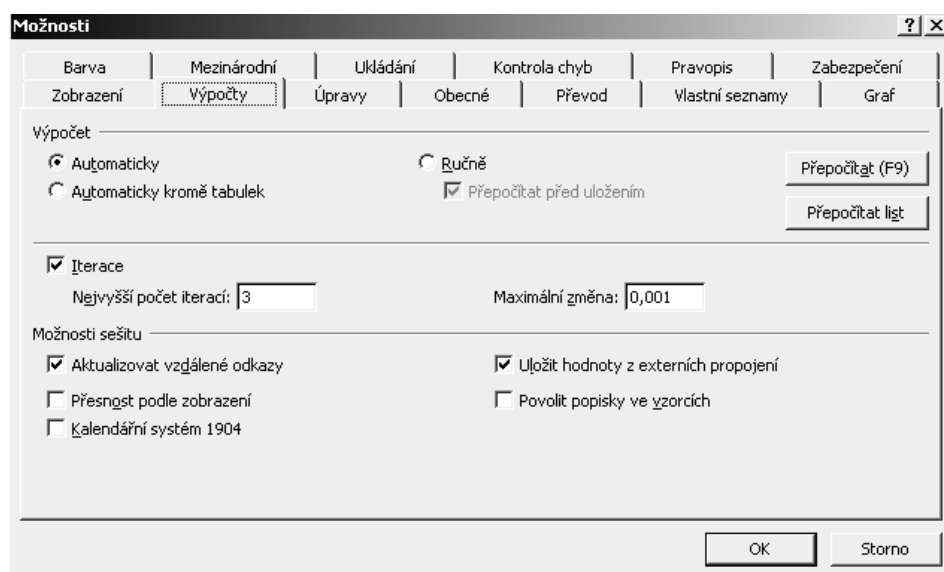
Když Evolver rozhodne o výrobě položky (ano/ne=1/0 sloupec ‘R’), pak se bude vyrábět (=1=ano). Jinak, když je nedostatek potřebného počtu výrobku i přes to, že je něco na skladě a například současně jistá část vyrobena, pak rovněž ‘ano’, bude se vyrábět.

Sloupec V představuje hodnotu, kterou je nutno ještě vyrobit, přesněji odečtení od výroby určené Evolverem a množstvím na skladě. Konkrétně má příkaz tvar:

$$=KDYŽ(U11-T11-P11<0;0;U11-T11-P11)$$

Když potřeba nepřevyšuje aktuální množství, pak není třeba vyrábět (nulový počet potřebných výrobků – 0), v opačném případě je nezbytné vyrobit minimálně právě onen nedostatek.

Protože zde jde o cyklické odkazy, vzájemně závislé podmínky, musíme v Excelu zvýšit počet iterací výpočtu na 3 iterace v nabídce ‘Nástroje/Možnosti...’, viz Obr. 44.



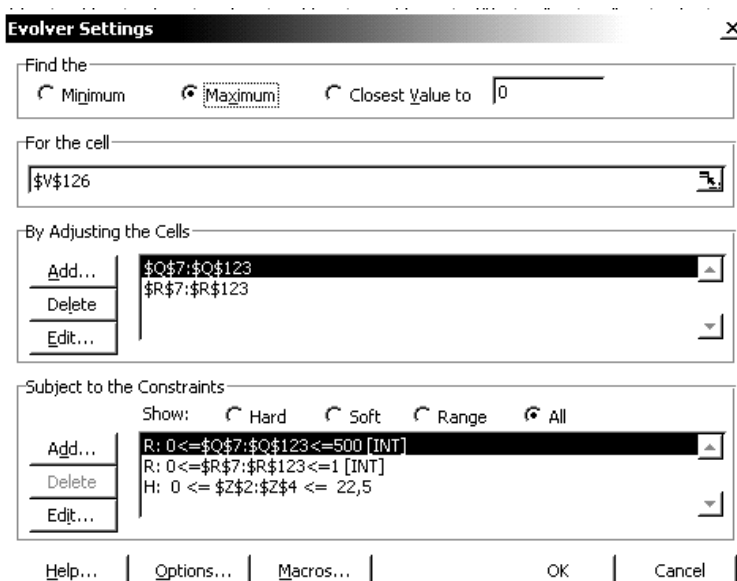
Obr. 44. Zvýšení počtu iterací v Excelu.

Sloupce 'W', 'X' a 'Y' obsahují násobky párů počtu vyrobených kusů na jednom ze tří druhů lisů pod označením jejich počtu 15, 5 a 1. Počet minut vulkanizace u každé položky je uveden ve sloupci 'C'. Sloupec 'AA' představuje množství na skladě na začátku po odečtení zakázek (sloupec 'U') od výroby a skladových zásob. Jednoduchý odpočet samozřejmě nestačí, v případě nesplnění potřebné výroby by hodnota byla záporná, což je zjevný nesmysl. Problém vyřešíme snadnou podmínkou, když je rozdíl záporný, napiš nulu. V 2. až 4. řádku ve sloupci 'U' a 'V' je součet hodin vulkanizací všech výrobků, jinak řečeno, čas provozů jednotlivých lisů. Toto vše vykonáme pro celý měsíc září 2004. Soboty a neděle mají pouze dvanáctihodinové směny.

Evolver, software pro řešení složitých problémů genetickými algoritmy

Nyní přecházíme na aplikaci genetických algoritmů. V kapitole 3.1.4 jsme vypočetli pravděpodobnou následující hodnotu u každého výrobku nejvhodnějším modelem, a tuto hodnotu vložíme do posledního sloupce, od kterého odečteme veškerou uvažovanou výrobu. Výsledné hodnotě můžeme určit důležitost prostým váhovým koeficientem, stejně tak u každého výrobku v každém měsíci. Ukázkou výpočtu uvedu pro druhý den, viz Obr. 43.

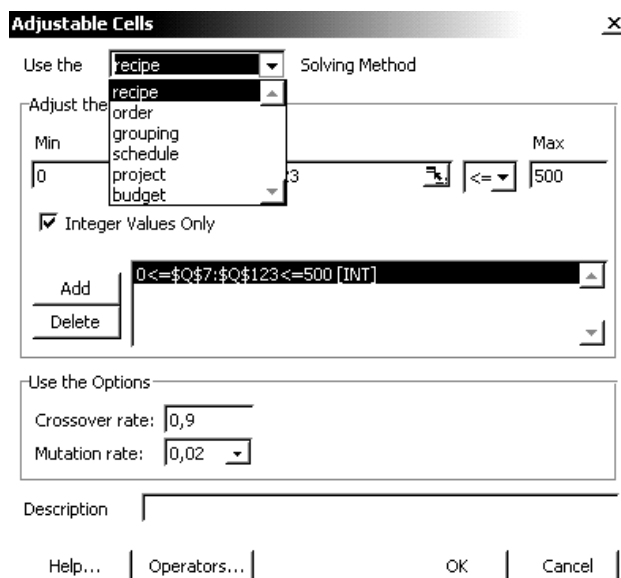
Vložení nezávisle proměnných a podmínek pro správný výpočet je velmi podobné s Řešitelem v Excelu, která má ale jen velmi omezené schopnosti při řešení složitých problémů, viz Obr. 45.



Obr. 45. Základní zobrazení Evolveru.

Do políčka 'For the cell' vložíme funkci, kterou chceme v průběhu optimalizace minimalizovat, maximalizovat nebo přiblížit vložené hodnotě, v našem případě se budeme snažit maximalizovat celkový výstup provozu, suma všech produktů. K tomu je ale nutné také určit buňky, které bude algoritmus upravovat pro dosažení nejlepší kombinace pro výstup a stanovit podmínky a omezení. Buňky pro úpravu se vkládají do políčka 'by adjusting the cells', viz Obr. 45., podmínky a omezení se vkládají přes políčka 'subject to the constraints'. Sloupec 'R' (vyrobit tento výrobek – ANO/NE) bude obsahovat pouze binární hodnoty, tedy omezení na celá čísla v rozsahu $\langle 0;1 \rangle$ a u sloupce 'Q' (velikost výroby), který označíme jako sloupec pro úpravu, nastavíme také celá čísla, ale s hranicí například $\langle 0;500 \rangle$ a metodu pro výpočet ('solving method') v obou případech 'recipe' viz Obr. 46. Program nám sice nabízí i jiné metody, ale pro naše účely je vhodná pouze tato. Vysvětlení ostatních jsou podle "helpu" následující:

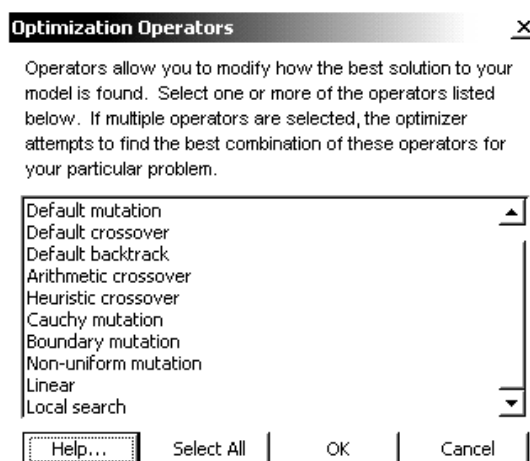
- Recipe solving method: všechny vložené hodnoty jsou pro řešení označena jako nezávislá a algoritmus optimalizuje jejich nejvhodnější skladbu, recepturu,
- Order solving method: vložené hodnoty představují pořadí kroků nějakého procesu a algoritmus určí jeho nejlepší složení,
- Grouping solving method: algoritmus předpokládá, že vložená data obsahují omezené množství určitých hodnot, neboli předpokládá skupiny představované danými hodnotami,
- Budget solving method: metoda předpokládá, že součet hodnot musí být vždy stejný,
- Project solving method: algoritmus pro plánování pořadí kroků v procesu s určením předcházejících kroků. Metoda je podobná s 'order solving method',
- Schedule solving method: algoritmus je určen pro úlohy typu vytvoření rozvrhu hodin.



Obr. 46. Nastavení parametrů nezávisle proměnné.

Mutační konstantu nastavíme na nízkou hodnotu 0,02, což je v praxi přibližně velmi častá hodnota, a práh křížení na hodnotu 0,9, která určuje pravděpodobnost, že se gen chromozomu nezmění, viz také Obr. 46. Mutační konstanta může být zvolena i 'auto', tzn. automatické zvolení nejlepší hodnoty a při konvergenci poměrné zvyšování, my však, jelikož známe povahu problému, budeme konstantu během procesu řešení upravovat ručně.

Software není omezen jen na pouhé základní algoritmy pro proces selekce-křížení-mutace, poskytuje i modifikované algoritmy vhodnější pro specifické problémy, viz Obr. 47.



Obr. 47. Modifikace algoritmů.

Přesněji, krom základních (default) algoritmů, jsou popsány zde:

- arithmetic crossover: aritmetická kombinace dvou rodičů,
- heuristic crossover: využívá pozitivního trendu změny jedinců,
- Cauchy mutation: malé změny při mutaci, občas větší,
- boundary mutation: pro rychlejší nalezení optima dává hraničním hodnotám vyšší váhy,
- non-uniform Mutation: čím je vyšší počet pokusů, tím snižuje mutační konstantu pro konečné doladění,
- linear: vhodné pro lineární problémy,
- local search: algoritmus se zaměřuje na nejbližší minimum funkce a dále doladuje.

Pokud zvolíme všechny uvedené algoritmy, to program dovoluje, vybírá při procesu řešení nejlepší možnost. Pro náš případ, který opravdu není složitý, vybrat vše můžeme. Z počátku Evolver pravděpodobně nezvolí Cauchyho mutaci, ale pro doladění a předčasnou konvergenci, je algoritmus vhodný.

Vyřešení problému na jeden den je ve velké většině vyřešen už po 2 minutách, a to i přesto, že byla zadána potřebná výroba využívající lisy z 90 % a nebyla algoritmu poskytnuta pomoc stanovit potřebnou výrobu již na začátku. To je dobrý výsledek.

3.3.2 Problém obchodního cestujícího

Přestože název může trochu zavést a nezdá se být následný kapitole 3.3.1, opak je pravdou. Problém obchodního cestujícího je sice označen jako NP problém, ale algoritmus je samozřejmě možné propůjčit i například problémům typu stanovení vhodného pořadí výrobků při penalizaci z přechodu mezi výrobky ve výrobě, kde například penalizace přechodu z výrobku A na výrobek B může být odlišná pro opačné pořadí (přechod z výrobku B na výrobek A). Koneckonců se toto specifikum může objevit i v realitě u problému obchodního cestujícího z důvodu jednosměrných ulic. Přestože zde půjde pouze o neukončený okruh, přesto mě tento detail neodradí od uvedeného názvu kapitoly. Vytvoření hamiltonovy kružnice při perturbaci permutací je totiž ve vybraném algoritmu velmi jednoduché, spočívá pouze v kratičkém doplnění algoritmu o přechod mezi prvním a posledním, taková úprava by

neměla trvat více než pár desítek sekund. Řízenou pertubaci permutací tedy použijí při řešení problému.

Pro stanovení pořadí výroby je nejprve nutné získat data z provozu, která obsahují čas přechodu z receptury na recepturu u stroje zvaného TROESTER, viz Obr. 48.

zavádění rozměrů 1.4. - 4.4. 2002 na VL TROESTER 200/75			
1.4.2002			
č.	změna rozměru	čas	doba (min)
1.	710 na 716	0:04	1
2.	716 na 723	1:34	2
3.	723 na 715	2:41	2
4.	715 na 712	3:18	2
5.	712 na 016	5:10	42
6.	016 na 299	6:32	21
7.	299 na 007	7:19	11
8.	007 na 047	8:10	11
9.	047 na 284	8:46	10
10.	284 na 046	9:30	20
11.	046 na 131	11:12	5
12.	131 na 335	11:22	0
13.	335 na 716	22:07	10
			10,5
2.4.2002			

Obr. 48. Přechodové časy mezi recepturami [29].

Bohužel, podnik má k dispozici jen asi 50 přechodů, a jelikož v září 04 bylo 117 produktů se 105 recepturami, není možné vytvořit kompletní řešení. Přesto pro stroj navrhnu konkrétní řešení určení pořadí zpracování druhů surovin (receptur), protože tento problém považuji za velmi potřebný k optimalizaci. Nejprve jsem si musel přechody, u kterých nebyly reálné informace, vygenerovat v Matlabu. Průměrná hodnota náhodných čísel byla 11,5 minut. Poté jsem se snažil, také v Matlabu, vytvořit algoritmus pro řešení problému obchodního cestujícího genetickými algoritmy, abychom potvrdili rámec kapitoly 3.3. Přesný postup popíši komentáři zdrojového kódu v následujícím pořadí, jak by mělo vypadat řešení:

1. Stanovení konstant

Především musím upozornit, že se nejedná o typický evoluční algoritmus. Chybí zde proces křížení, a tedy i počáteční selekce. Zastoupení má pouze mutace a následný výběr lepšího jedince. Potřebné konstanty jsou:

- 1 pop=100; - velikost populace
- 2 fmk=0.1; - mutační konstanta F
- 3 pg=20000; - počet generací

2. Generace jedinců

Za minimálně účinné množství chromozomů, jedinců, se považuje dvanáct (dvanáctičlenná populace), přesto pro jistotu zvolím populaci o sto jedincích. Receptury jsem označil čísly od 1 do 105 a vytvořil jsem sto náhodných permutací 105 receptur pomocí funkce ‘randperm’:

```
4 for a=1:pop
5     jedin(:,a)=randperm(i);
6 end
```

3. Křížení

Přestože tento krok ve výpočtu přeskakují, považuji jej za hodný uvedení. V našem specifickém problému se jedná o přenesení náhodné části jedince na druhý. Protože v případě přenesení pouze jednoho genu se jedná v podstatě o krok 4., mutaci, je dobré, aby náhodná část byla posloupnost více hodnot chromozomu. Křížení je velmi dobré pro přenášení nejlepších vlastností na jiné jedince.

4. Mutace

Literatura nabízí dva základní druhy mutace, a to ‘exchange mutation’ a ‘shift mutation’. Jak názvy napovídají, v prvním případě se jedná o výměnu libovolného počtu genů v chromozomu a v druhém případě jde o vyjmutí zvolených genů a umístění na libovolné pořadí v chromozomu. Zvolil jsem ‘shift mutation’, z důvodu menšího “rozházení” při novém přeskupení. Zatímco v posuvné mutaci při přesunutí jednoho genu dojde ke třem změnám, ve výměnné mutaci dojde ke čtyřem. Přesněji následující zdrojový kód:

```
7 psm=ceil(round(rand(1,1)*(i-1)+1)*fmk)
8 hpm=randperm(i).';
9 hpm=hpm(1:2*psm,1);
```

Zde ‘psm’ představuje náhodný počet posuvných změn do maximální výše ‘fmk’ procent. ‘hpm’ pak představuje náhodný výběr konkrétních receptur maximálně do výše ‘psm’.

```
10 for b=1:psm
11     if hpm(b)<hpm(psm+b-1)
12         if hpm(b)==1 & hpm(psm+b-1)==i
13             elseif hpm(psm+b-1)==i
14                 elseif hpm(b)==1
15                     else
```

```
16         end
17     elseif hpm(b)>hpm(psm+b-1)
18         if hpm(b)==i & hpm(psm+b-1)==1
19             elseif hpm(psm+b-1)==1
20                 elseif hpm(b)==i
21                     else
22                         end
23         end
24     end
```

Nyní již mutace samotná. Protože jde o ‘psm’ změn, pak počet cyklů musí být stejný, viz řádek 10. Mutace má jednoduchý princip, ale k naprogramování je třeba uvážit různé alternativy, a to pokud se gen přesouvá z předních či zadních pozic, viz řádek 11 a 17. Dále pak pokud je gen ze začátku řady a následné umístění na konci, nebo je gen na konci, nebo pokud je nové umístění na začátku řady, viz řádek 12 až 16. Pro případ kdy bude gen posunut nazpět, jsou podmínky inverzní, viz řádek 18 až 23 (všimněte si, že ‘==i’ a hodnota ‘==1’ jsou prohozeny). Pro lichý počet změn v případě záměny prostřední hodnoty za prostřední je v algoritmu úprava automaticky vyloučena. ‘&’ znamená funkci “a zároveň”. Naprogramování pro neprogramátora tedy nebylo triviální.

5. Výběr lepšího jedince

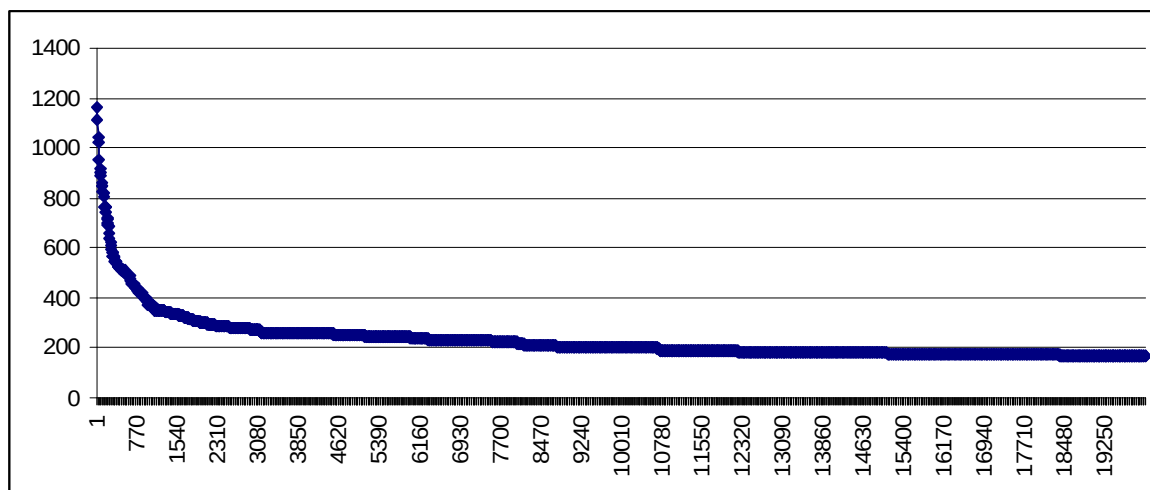
Výběr lepšího jedince je ve své podstatě obecně načtení celé funkce pro výpočet konečné hodnoty. Obvykle to bývá na čas nejnáročnější operace, v našem případě to však není pravda, viz zdrojový kód:

```
25 for b=1:i-1
26     jedinvsm=jedinvsm+parypp(jedinsm(b),jedinsm(b+1));
27 end
```

Chybí sice cyklus pro celou populaci, ale to je otázka dvou řádků. Matice ‘parypp’ představuje kombinační tabulku časů přechodů mezi všemi páry receptur, neboli vzdálenosti pro všechny možné páry měst, pro lepší pochopení.

6. Konečné řešení

Pro přijatelné řešení stačilo pouhých 10000 generací, to představuje pět minut na středně výkonném počítači. Já jsem zvolil generací 20000, výsledek viz Obr. 49.



Obr. 49. Průběh nelepších hodnot při řešení problému obchodního cestujícího.

Výsledná hodnota 167 minut ztrát při výměně 105 receptur, tzn. 104 změn, je s porovnáním s počáteční náhodnou permutací pořadí receptur (1162 minut) velmi dobrý výsledek. Pro bližší představu původní průměrný čas na přenastavení TROESTERU byl $1162/104 = 11,17$ minut. V nově uspořádaném pořadí jsme během desíti minut dospěli k potřebě pouhé 1,61 minuty! na přenastavení stroje. Výsledek představuje 7x lepší využití času, což při velmi častých změnách může v takové společnosti jako Mitas, a.s. představovat statisícové částky ročních úspor v nákladech na zaměstnance, protože mohou ušetřený čas upotřebit na jiné aktivity, které by vykonávali jiní pracovníci

Volba mutační konstanty se ukázala jako ne příliš významná, jen je důležité, aby nenabývaly příliš vysokých hodnot, nejlépe do 0,3. Nastavení mutační konstanty na hodnotu 50 % a výše není možné, to vychází zejména z povahy algoritmu, který problém neumí řešit. Z praktického hlediska ani nemá smysl algoritmus upravit, protože výsledek po výměně více jak 50 % hodnot bude velmi pravděpodobně mít podobný úspěch jako vytvoření nového náhodně složeného jedince. Vyřešení problému by mělo značně malý mezní užitek. Výměna méně hodnot bude jistě efektivnější zejména v konečném doladění.

Samozřejmě je třeba uvažovat průběžně vysoké využití kapacit lisů, pak je na místě přemýšlet o složitější funkci, která by zároveň průběžně maximalizovala využití kapacit například každých 30 minut. Postoj společnosti se bohužel změnil.

3.4 Implementace statistických metod do procesu plánování výroby

Snad nejvhodnější volba bez dlouhého přemýšlení je zaměstnat pracovníka ekonomicky vzdělaného s alespoň základními znalostmi statistiky, který zvládá rovněž základy programování v nejběžnějších programovacích jazycích. Člověk, který by tuto relativně všeobecnou práci dovedl k detailnímu zpracování řešící položený problém, s motivací, ochotou stále se učit a schopností nejen přijmout položenou výzvu, jejíž konečný cíl by jistě splnil veškerá osobní očekávání, je pochopitelně velmi vzácný. Nalezení vhodného uchazeče zřejmě vyžaduje nemalé úsilí.

Alespoň základní zaškolení pracovníka s nástupním zaslouženým platem blízkým mediánu mzdy v ČR (16 500 Kč) by trvalo přibližně šest měsíců. Následný osobní sběr potřebných informací rovněž šest měsíců a vytvoření funkčního modelu plánování výroby čtyři měsíce. Nevylučuji, samozřejmě, řešení dílčích problémů nejen dle aktuálního požadavku (například vytvoření modelu pro optimální lokální pohyb při přesunu materiálu). Požadavky na sběr informací jistě překročí rámec schopností jedné osoby, uvažujme proto využití i jiných pracovníků v průměru na 20 minut denně včetně nepracovních dnů.

Software Matlab mohu jen doporučit, jen s cenou to není nejlepší. Cena jednouživatelská licence činí 87 000 Kč, avšak při spolupráci s Univerzitou Tomáše Bati, FAME, nejlépe s doktorandem, klesá na nulovou hodnotu, UTB výpočetní software Matlab (aplikovaný také v NASA) vlastní.

Celkové náklady k uvedení modelu do praxe mohou být “jen” $16500\text{Kč} \times (24 + 4)\text{měs.} + 70\text{Kč/hod} \times 0,3\text{hod} \times 365 \times (2 + 1/3)\text{roků} \approx 480000\text{Kč}$ za rok a půl včetně běžného platu pracovníka. V případě pokračující spolupráce s doktorandem, kdy bude nutné SW Matlab zakoupit, narostou náklady relativně pouze o 18 %.

ZÁVĚR

Zřejmě nejdůležitějším poznatkem při zpracování práce tohoto typu je nutnost neustále rozvíjet své znalosti nejen předmětu Statistika. Mezery v komplexní a hodnotné aplikaci statistických metod jsou častým jevem a zřejmě jsem se tomu nevyhnul ani já již z důvodu naprosté nekompetence být pro jednodušskou literaturu adekvátně kritický, nejsem statistik. Přestože pokus o osvojení základních metod vhodných pro dané téma vyžadoval nemalé úsilí a trpělivost a považuji jej za relativně dostačující.

Téma diplomové práce bylo dobrým prostředníkem k dávnému nerealizovanému přání zvládnutí v základu dostačující problematiku časových řad obsahující ne příliš statistický nástroj umělé neuronové sítě. Ostatní metody, zdánlivě méně obtížné, mi rovněž poskytly uspokojení ve vlastní praktické aplikaci. Harmonická analýza se ukázala jako snadnější, odtud volba vlastního naprogramování celého algoritmu s příslušnými statistikami, naopak Box-Jenkinsonova metodologie se projevila jako obtížná ve svém konečném důsledku, který vyžaduje složitější postupy nejen ve fázi analýzy a úpravy vstupních dat. SARIMA modelování je velmi nepraktické (nevhodné pro efektivní aplikaci), vyžaduje časté grafické zhodnocení v rámci fáze volby správných autokorelačních funkcí. Jistě existuje speciální software a různá složitá rozšíření v rámci Matlabu a jiných častěji užívaných programů řešící i tento problém, ale vynaložené úsilí v rámci diplomové práce považuji za velmi neproduktivní.

Nejvíce si však cením zvládnutí potřebného rozsahu nástrojů velmi předimenzovaného pomocníka pro náš úkol, a to Matlabu. Cením si jej zejména proto, že mi otevřel cestu k samostatnému řešení téměř jakéhokoliv vlastního problému, který by se jistě mohl setkat (a samozřejmě v práci setkal) s omezením při řešení problému v běžně dostupných aplikacích (jako je například velmi rozšířený nástroj tabulkový procesor Excel od společnosti Microsoft). To, co by trvalo v Excelu velmi zdlouhavě, například přenášení vzorců, může být v Matlabu pomocí cyklů vyřešeno jen ve čtyřech řádcích zdrojového kódu. Pro snadnější výpočty s obsahem méně závislostí je však Excel velmi dobrou volbou. Pochopitelně vše má své klady a vše má své zápory.

Genetické algoritmy se ukázaly jako velmi dobré řešení k problému obchodního cestujícího a k 'intuitivnímu' rozplánování výroby. Vlastní řízenou pertubací může nahradit lidskou schopnost a nutnost opakované analýzy podmínek nezbytných k vypracování plánu výroby.

Nemohu zapomenout na konkrétní popis řešení. Můj postup byl tedy vytvoření predikce prodeje, popis výroby a konečné rozplánování v závislosti na předpovědi a znalostech o výrobě. Byl jsem seznámen s hrubým procesem výroby a podle konzultací s pracovníky oddělení průmyslového inženýrství vlastním poznatkem o pochopitelné složitosti jednotlivých procesů výroby motoplášťů. Skutečná aplikace statistických i nestatistických metod do veškerých podrobností bude vyžadovat značné úsilí.

SEZNAM POUŽITÉ LITERATURY

- [1] ARLT, J. Moderní metody modelování ekonomických časových řad. Praha: Grada Publishing, 1999. 312 s. ISBN 80-7169-539-4.
- [2] ARLT, J.; Kozák, J.; Hindls, R. Úvod do analýzy ekonomických časových řad. Praha: Vysoká škola ekonomická v Praze, 1994. 208 s. ISBN 80-7079-760-6.
- [3] CIPRA, T. Analýza časových řad s aplikacemi v ekonomii. Praha: SNTL, 1986. 248 s. ISBN 01-012-86.
- [4] CYHELSKÝ, L. Úvod do teorie popisné statistiky. Praha: SNTL, 1974. 372 s. Bez ISBN.
- [5] CYHELSKÝ, L.; KALOUNOVÁ, J.; HINDLS, R. Elementární statistická analýza. Praha: Victoria Publishing, 1996. 320 s. ISBN 80-85943-18-2.
- [6] ČERMÁK, V. Statistika. Praha: SNTL, 1968. 205 s. ISBN 04-302-68.
- [7] DOSTÁL, P. Moderní metody ekonomických analýz: finanční kybernetika: Univerzita Tomáše Bati, Fakulta managementu a ekonomiky, 2002. 110 s. ISBN 80-7138-075-8.
- [8] FABIAN, V. Základní statistické metody. Praha: Nakladatelství Československé akademie věd, 1963. 451 s. ISBN 21-135-63.
- [9] FANTA, J. Neuronové sítě ve společenských vědách. Praha: Karolinum, 2000. 165 s. ISBN 80-246-0175-3.
- [10] HAKL, F. Úvod do teorie neuronových sítí. Praha: Vydavatelství ČVUT, 1998. 210 s. ISBN 80-246-0175-3.
- [11] HINDLS, R.; HRONOVÁ, S. Statistika v hospodářství. Praha: ETC Publishing, 1998. 636 s. ISBN 80-86006-56-5.
- [12] HINDLS, R.; HRONOVÁ, S.; NOVÁK, I. Metody statistické analýzy pro ekonomy. Praha: Management press, 2000. 249 s. ISBN 80-7261-013-9.
- [13] KLÍMEK, P. Ekonometrie. Zlín: Univerzita Tomáše Bati ve Zlíně, 2001. 72 s. ISBN 80-7318-004-9.

- [14] MARČEK, D. Neuronové sítě a fuzzy časové řady. Opava: Slezská univerzita v Opavě, Filozoficko-přírodovědecká fakulta, 2002. 151 s. ISBN 80-7248-157-6.
- [15] NOVÁ, M. Umělé neuronové sítě: teorie a aplikace. Praha: C.H. Beck, 1998. 382 s. ISBN 80-7179-132-6.
- [16] PAVELKA, F.; KLÍMEK, P. Aplikovaná statistika. Zlín: Vysoké učení technické v Brně, Fakulta managementu a ekonomiky ve Zlíně, 2000. 131 s. ISBN 80-214-1545-2.
- [17] PAVELKA, F.; RYTÍŘ, V. Metody statistické analýzy. Zlín: Vysoké učení technické v Brně, Fakulta managementu a ekonomiky ve Zlíně, 1999. 118 s. ISBN 80-214-1503-7.
- [18] KAŇOKOVÁ, J. Teorie statistiky pro řízení a plánován. Praha: SNTL, 1989. 400 s. ISBN 04-301-89.
- [19] LIKEŠ, J. LAGA, J., Základní statistické tabulky. Praha: SNTL, 1978. 491 s. ISBN 01-339-78.
- [20] MÜLLER, T. Interaktivní tvorba rozvrhu. Praha: Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, 2001. 77 s. bez ISBN.
- [21] RAO, R.; RAO, C. Lineární metody statistické indukce a jejich aplikace. Praha: Nakladatelství Československé akademie věd, 1978. 666 s. ISBN 104-21-852.
- [22] SEGER, J. Statistické metody pro ekonomy průmyslu. Praha: SNTL, 1988. ISBN 04-313-88.
- [23] SEGER, J.; KOZÁK, J. Jednoduché statistické metody v prognostice. Praha: SNTL, 1975. ISBN 04-328-74.
- [24] ŠEDA, M. Teorie grafů. Brno: Vysoké učení technické v Brně, 2003. Bez ISBN. http://www.fme.vutbr.cz/kam/opory/pdf/uai/teorie_grafu/TG03.pdf.
- [25] ŠÍMA, J. Neuronové sítě a neuropočítače. Praha: Vydavatelství ČVUT, 1998. 124 s. ISBN 80-01-01455-X.
- [26] ŠNOREK, M., R. Teoretické otázky neuronových sítí. Praha : Matfyzpress, 1996. 390 s. ISBN 80-85863-18-9.

- [27] WONNACOT, T., H; WONNACOT, R., J. Statistika pro obchod a hospodářství. Praha: Victoria Publishing, 1992. 891 s. ISBN 80-85605-09-0.
- [28] ZELINKA, I. Umělá inteligence v problémech globální optimalizace. Praha, BEN – technická literatura, 2002, 192. s. ISBN 80-7300-069-5.
- [29] Interní materiály společnosti Mitas, a.s.
- [30] Umělé neuronové sítě a jejich použití v lékařských datech,
<http://www.uivt.cas.cz/~roman/pub/tisk/automat1.doc>.
- [31] <http://www.mitas.cz/>

SEZNAM OBRÁZKŮ

Obr. 1. ACF.	15
Obr. 2. PACF.	16
Obr. 3. Matematický model neuronu.	18
Obr. 4. Vícevrstvá síť.	19
Obr. 5. Standardní sigmoida.	19
Obr. 6. Dvou-bodové křížení.	28
Obr. 7. Shift mutation.	28
Obr. 8. Flowchart provozu na výrobu motoplášťů.	33
Obr. 9. Výřez z ročního harmonogramu činnosti provozovny na výrobu motoplášťů.	36
Obr. 10. Matice vyhlazených dat pomocí prostého centrovaného klouzavého průměru.	38
Obr. 11. Matice vyhlazených dat pomocí prostého centrovaného klouzavého průměru.	39
Obr. 12. Původní neočištěné hodnoty.	39
Obr. 13. M.S.E.	41
Obr. 14. Reálný prodej a teoretické hodnoty výrobku '21212'.	42
Obr. 15. Očištěná historická data výrobku '27393' o trend.	45
Obr. 16. Proložení skutečných hodnot soustavou sinových a kosinových křivek produktu '27393' očištěných o trend.	46
Obr. 17. Periodogram výrobku '27393'.	47
Obr. 18. Složená periodicitu 8. a 6. periody u výrobku s id '27393'.	49
Obr. 19. Rozdíl v reziduích před a po aplikaci DW statistiky.	51
Obr. 20. Matice výsledků po všech potřebných výpočtech.	52
Obr. 21. Prodej výrobku '26468'.	54
Obr. 22. Stacionarizovaná řada hodnot z Obr. 21 řádem 1.	55
Obr. 23. $l = 0,5$	56
Obr. 24. $l = 1$	56
Obr. 25. $l = -1$	56
Obr. 26. $l = -0,5$	56
Obr. 27. $l = 0$	56
Obr. 28. ACF.	57
Obr. 29. PACF.	57

Obr. 30. Výsledek funkce SARIMA (1,1,0)(1,0,0) v SW Statistica 6.	58
Obr. 31. Kontrola nové tabulky PACF.	59
Obr. 32. Predikce funkce SARIMA (1,1,0)(1,0,0) v SW Statistica 6.	59
Obr. 33. Reálné hodnoty a SARIMA (1,1,0)(1,0,0).	60
Obr. 34. Zvolená topologie neuronové sítě.	62
Obr. 35. Train genetic.	63
Obr. 36. Advanced Genetic Options.	63
Obr. 37. Model neuronové sítě v Neurosolutions 4.32.	64
Obr. 38. Parametr b přenosové funkce.	64
Obr. 39 Chyby testovací a validační množiny dat.	64
Obr. 40. Proložení dat nástrojem ANN.	65
Obr. 41. Excel – ANOVA s opakováním.	67
Obr. 42. Nástroj Regrese v Excelu.	69
Obr. 43. Výřez z Excelu.	71
Obr. 44. Zvýšení počtu iterací v Excelu.	72
Obr. 45. Základní zobrazení Evolveru.	73
Obr. 46. Nastavení parametrů nezávisle proměnné.	75
Obr. 47. Modifikace algoritmů.	75
Obr. 48. Přechodové časy mezi recepturami.	77
Obr. 49. Průběh nelepších hodnot při řešení problému obchodního cestujícího.	80

SEZNAM TABULEK

Tab. 1. Periodogram.....	14
Tab. 2. Hodnoty l s odpovídajícími transformacemi	17
Tab. 3. ANOVA tabulka pro dva faktory s opakováním.....	24
Tab. 4. ANOVA tabulka pro dva faktory s opakováním, statistika F.	24
Tab. 5. Přehled lisů v provozu.	34
Tab. 6. Reálné objednávky, reálný prodej a rezidua vypočtené pomocí teoretických hodnot.....	40
Tab. 7. Vypočtené teoretické hodnoty pro periodu osmi měsíců	45
Tab. 8. Fisherův a Sieglův test periodicity.....	48
Tab. 9. Výřez výchozí matice pro následné řešení.	49
Tab. 10. Tabulka pro nástroj ANOVA.....	68
Tab. 11. Výstup nástroje ANOVA v Excelu.....	68
Tab. 12. Výstup 1.....	69
Tab. 13. Výstup 2.....	69
Tab. 14. Výstup pro hodnocení koeficientů.....	70

SEZNAM PŘÍLOH

P I Princip genetického algoritmu zvaného diferenciální evoluce.

PŘÍLOHA P I: PRINCIP GENETICKÉHO ALGORITMU ZVANÉHO DIFERENCIÁLNÍ EVOLUCE.

