

Vedecká rada Stavebnej fakulty  
Slovenskej technickej univerzity v Bratislave

---

Ing. Tomáš Bacigál

Autoreferát dizertačnej práce

**Moderné metódy modelovania  
viacrozmerných časových radov  
s aplikáciou v geodézii**

na získanie vedecko–akademickej hodnosti  
*philosophiae doctor*  
v odbore doktorandského štúdia  
11-14-9 Aplikovaná matematika

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Katedre matematiky a deskriptívnej geometrie SvF STU v Bratislave.

- Predkladateľ:** Ing. Tomáš Bacigál  
Katedra matematiky a deskriptívnej geometrie SvF STU  
Radlinského 11  
813 68 Bratislava
- Školiteľ:** Prof. RNDr. Magda Komorníková, CSc.  
Katedra matematiky a deskriptívnej geometrie SvF STU
- Oponenti:** Ing. Juraj Janák, PhD.  
Katedra geodetických základov SvF STU  
Radlinského 11  
813 68 Bratislava
- Prof. RNDr. Milan Mareš, DrSc.  
Ústav teórie informácie a automatizácie  
AV ČR  
Pod Vodárenskou věží 4  
182 08 Praha
- Doc. RNDr. Karol Pastor, CSc.  
Kat. apl. mat. a štat. FMFI UK  
Mlynská dolina  
842 48 Bratislava

Autoreferát bol rozoslaný dňa ..... 2007

Obhajoba dizertačnej práce sa uskutoční dňa ..... 2007 o ..... hod. na Stavebnej fakulte STU, Radlinského 11, 813 68 Bratislava, v seminárnej miestnosti KMaDG, pred komisiou pre obhajobu dizertačnej práce doktorandského štúdia vymenovanou dňa ..... 2007 predsedom spoločnej odborovej komisie vo vednom odbore 11-14-9 Aplikovaná matematika.

**Predseda spoločnej odborovej komisie:**  
Prof. RNDr. Zdenka Riečanová, CSc.  
Katedra matematiky FEI STU  
Ilkovičova 3  
812 19 Bratislava

# Obsah

Úvod	2
1 Súčasný stav	2
2 Ciele dizertačnej práce	7
3 Dosiiahnuté výsledky	8
3.1 Viacrežimové modely so spojitým prechodom . . . . .	8
3.2 Spoločný trend a sezónna zložka . . . . .	11
3.3 Modelovanie závislosti pomocou kopúl . . . . .	12
3.4 Výstavba modelu s programom v systéme Mathematica . . . . .	12
Perspektívy ďalšieho výskumu	15
Zoznam publikácií a citácií	16
Literatúra	17
Summary	19

# Úvod

Technické disciplíny v stavebníctve ako napr. geodézia, geológia, statika budov a pod. pracujú s geometrickými a fyzikálnymi veličinami a skúmajú tak procesy, ktoré vplyvajú na naše prirodzené aj umelo dotvorené prostredie. Zdokonaľovanie meracích prístrojov a automatizácia zberu údajov umožňujú stále detailnejší monitoring týchto procesov, pochopiteľne, ruka v ruke s efektívnymi metódami spracovania.

V geodézii je jedným z hlavných záujmov presné určenie polohy bodu (nielen) na zemskom povrchu a pre tento cieľ existuje veľké množstvo prístrojov a techník, ktorými sa realizujú všetky potrebné sprostredkujúce veličiny. V súčasnosti možno za najefektívnejší nástroj na určenie polohy považovať systémy využívajúce obieh umelých družíc okolo Zeme, menovite americký GPS, ruský GLONAS, či európsky projekt Galileo, ktoré pri aplikácii špeciálnych geodetických techník dosahujú až milimetrovú presnosť. Pre zvýšenie presnosti v komerčnej praxi i pre výskumné účely sú na referenčných bodoch zriadené trvalé observácie. Výsledné realizácie polohy (vektorová náhodná premenná v troch priestorových zložkách) sú vo forme časového radu cenným zdrojom informácií o procesoch podieľajúcich sa na dynamike jedného bodu alebo/a celého územia.

Analýza časových radov má v štatistike pomerne silnú podporu, najmä vďaka aplikáciám v ekonómii a finančníctve, pre ktoré väčšina doteraz všeobecne používaných matematických modelov a techník vznikla. Najpopulárnejšou modelovou štruktúrou pre popis sociálno-ekonomických procesov je autoregresia, teda odvodenie aktuálneho stavu daného procesu od jeho minulých efektov. Táto myšlienka sa ukázala ako veľmi výhodná aj pri predpovediach do budúcnosti a v predkladanej práci tvorí základ pre uvedenie ďalších vylepšení. Na druhej strane a v kontraste s týmto prístupom stojí hľadanie vzťahov medzi súčasnými prejavmi skúmaného procesu vo viacerých premenných s cieľom pochopiť interakcie v určitom systéme (príroda, ľudská spoločnosť). Perfektným príkladom modelovania vzťahov sú tzv. kopule. V praxi sa však čoraz častejšie začína vyskytovať otázka individuálnej dynamiky a vzájomných vzťahov súčasne, preto aj naša práca sa nesie v duchu modelovania viacrozmerných náhodných premenných (ďalej tiež označované ako viacrozmerné modelovanie), ktoré dokáže lepšie postihnúť vlastnosti reálneho sveta ako doteraz používané jednorozmerné modely.

## 1 Súčasný stav

Pod pojmom časový rad  $\{y_1, y_2, \dots, y_n\}$  rozumieme sekvenciu údajov usporiadaných vzostupne podľa ich časového indexu, získaných pri pozorovaní určitého javu ako realizáciu stochastického procesu či už cez priamo merané alebo odvodené náhodné veličiny, zvyčajne v konštantnom časovom rozstupe. Zjednodušený zápis je  $y_t$ , pre  $t = 1, 2, \dots, n$ , kde  $t$  je časový index a  $n$  dĺžka časového radu. Hlavným zmyslom analýzy časových radov je pochopiť mechanizmus, ktorý generuje merané údaje, a predpovedať hodnoty v čase, pre ktorý merania nie sú dostupné.

V súčasnej, nielen geodetickej praxi je stále rozšírené modelovanie jednorozmerných časových radov a to pomocou rozkladu do štyroch základných zložiek [4]:

- trend, odráža dlhodobé zmeny v priemernom chovaní časového radu, napr.

vplyvom systematických zložiek pôsobiacich v jednom smere,

- sezónna zložka, popisuje periodické zmeny, ktoré sa odohrávajú behom jedného dňa, roka alebo inej prirodzenej periódy,
- cyklická zložka, menej pravidelné a v čase sa meniace fluktuácie okolo trendu,
- reziduálna zložka, je tvorená náhodnými výkyvmi bez rozpoznateľného systematického charakteru.

Reziduá v sebe nesú množstvo informácií, a preto sa naďalej analyzujú, najčastejšie pomocou tzv. Box-Jenkins metodológie, ktorá zahŕňa širokú skupinu lineárnych modelov počnúc autoregresným modelom rádu  $p$ , skrátene AR( $p$ ),

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (1)$$

kde  $\phi_i$ ,  $i = 1, \dots, p$ , sú neznáme parametre a  $\varepsilon_t$  predstavuje rad nekorelovaných náhodných veličín s rovnakým rozdelením pravdepodobnosti s konštantnou strednou hodnotou (najčastejšie nulovou) a s konštantným rozptylom  $\sigma_\varepsilon^2$ , tzv. biely šum. Z ďalších lineárnych modelov sa zvyknú aplikovať model kĺzavých priemerov (moving average) MA, zmiešaný model ARMA a model nestacionárnych časových radov ARIMA (pozri napr. [1][2][7]). Odhad parametrov je otázka bežnej metódy najmenších štvorcov (MNŠ) a v súčasnosti existujú mnohé softvéry, ktoré (buď ako samostatné výpočtové systémy, alebo doplnkové balíky pre ne) umožňujú vybudovanie modelu od predbežnej diagnostiky, cez určenie rádu, odhad až po testovanie reziduá a predpovede. Za všetky spomeňme balík *Time Series* pre komerčný systém *Mathematica*, alebo *stats* pre podobne zamerané, ale voľne šíriteľné programovacie prostredie *R*. Uvedené softvéry zároveň zvládajú aj viacrozmerné mutácie lineárnych modelov, napr. vektorový autoregresný model rádu  $p$  VAR( $p$ ) definovaný vzťahom

$$\mathbf{y}_t = \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \dots + \Phi_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (2)$$

kde  $\mathbf{y}_t = (y_{1t}, \dots, y_{kt})$  je  $k$ -rozmerná vektorová náhodná premenná,  $\Phi_i$  je  $k \times k$  matica neznámych parametrov a  $\boldsymbol{\varepsilon}_t$  je vektor jednotlivo nekorelovaných (ani s vlastnou ani s históriou ostatných) procesov bieleho šumu. Ak sú hodnoty  $\mathbf{y}$  v čase  $t$  ovplyvňované nielen predchádzajúcimi hodnotami  $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}$  ale aj tzv. exogénnymi vysvetľujúcimi premennými, povedzme  $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{lt})$ , kde  $l$  je ich počet, model skúmanej veličiny  $\mathbf{y}_t$  v konkrétnom čase  $t$  je postavený na jej vzťahu k rovnako aktuálnym hodnotám  $\mathbf{x}_t$ . Príkladom môže byť regresia závislosti deformácie stavebného objektu od teploty, prípadne vlhkosti vzduchu, objemu dopravy v jeho blízkosti a ďalších exogénnych premenných stochastickej povahy alebo aj priamo od niektorej deterministickej veličiny, najčastejšie času. Pre úplnosť dodajme, že nie je nereálny ani všeobecnejší VAR model, kde do regresie vstupujú súčasné a/alebo minulé hodnoty modelovaných aj vysvetľujúcich premenných, ak takýto model má reálny podklad a vieme ho rozumne interpretovať.

Pri viacrozmernom modelovaní nestacionárnych časových radov, zvlášť zavedením vysvetľujúcej premennej vzniká riziko tzv. falošnej alebo zdanlivej regresnej závislosti, kedy dve alebo viac premenných má veľmi podobný priebeh v dlhodobom

horizonte a koeficient determinácie regresie jednej od ostatných hovorí v prospech závislosti, avšak tým ešte nepotvrzuje aj ich krátkodobé previazania. Znamená to, že ich nemusí generovať ten istý proces. V súvislosti s lineárnymi modelmi sa vynára i ďalšia otázka, a to či modelované závislosti sú vôbec lineárne alebo proces generujúci pozorované údaje neprebíha v určitých režimoch, ako napr. voda, ktorej prejavy sú prítomné takmer všade, a ktorej vlastnosti úzko súvisia s jej skupenstvom, teda sa skokovo menia s teplotou. Napokon, hoci klasický regresný prístup je široko využiteľný, limituje analýzu klasifikovaním veličín na modelované (závislé) a vysvetľujúce (nezávislé) premenné, zatiaľčo náš zámer môže byť všeobecnejší a sústreďovať sa na pochopenie rozdelenia v systéme observovaných veličín, na združené rozdelenie pravdepodobnosti.

## Modely s premenlivými režimami

Viacrežimový autoregresný model patrí do skupiny nelineárnych, presnejšie pokúskoch-lineárnych modelov, ktoré vynikajú dobrou interpretovateľnosťou a možnosťou predikcie. Prechod medzi režimami je stochastický, riadený reláciou tzv. indikačnej premennej  $z_t$ , ktorá je pozorovateľná, a prahu  $r$ . Typický  $k$ -rozmerný 2-režimový model so skokovým prechodom – TAR model – v maticovom zápise má tvar

$$\mathbf{y}_t = \Phi_1 \mathbf{X}_t 1_{(z_t \leq r)} + \Phi_2 \mathbf{X}_t 1_{(z_t > r)} + \varepsilon_t, \quad (3)$$

kde  $\mathbf{y}_t$  je  $k \times 1$  modelovaný vektor,  $\mathbf{X}_t$  je  $K$ -rozmerný regresný vektor,  $\Phi_j$ ,  $j = 1, 2$ , je  $k \times K$  matica neznámych parametrov  $j$ -teho režimu, a funkcia  $1_A$  nadobúda hodnotu 1, ak udalosť  $A$  nastane, v opačnom prípade je rovná nule. Na druhej strane  $k$ -rozmerný 2-režimový model s hladkým (alebo spojitým) prechodom – STAR model – má tvar

$$\mathbf{y}_t = \Phi_1 \mathbf{X}_t (1 - G(z_t, \gamma, r)) + \Phi_2 \mathbf{X}_t G(z_t, \gamma, r) + \varepsilon_t, \quad (4)$$

kde  $G(z_t, \gamma, r)$  je aspoň dva razy diferencovateľná prechodová funkcia. Parameter  $\gamma > 0$  určuje "hladkosť" prechodu (pre  $\gamma \rightarrow \infty$  je prechod skokový, pre  $\gamma \rightarrow 0$  je prechod plynulý). Na Obr.1 sú zobrazené (pre rôzne hodnoty  $\gamma$  a  $r = 0$ ) dve najčastejšie používané prechodové funkcie, logistická prvého rádu a exponenciálna,

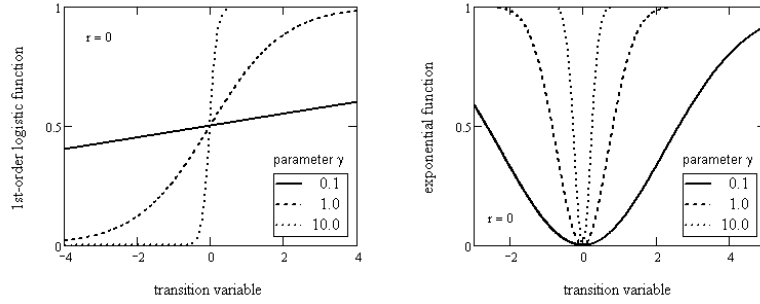
$$G(z_t, \gamma, r) = \frac{1}{1 + e^{-\gamma(z_t - r)}}, \quad G(z_t, \gamma, r) = 1 - e^{-\gamma(z_t - r)^2}, \quad \gamma > 0. \quad (5)$$

Regresný vektor obsahuje konštantný člen a všetky uvažované časovo posunuté endogénne (modelované) a exogénne (vysvetľujúce) veličiny (s možným časovým posunutím). Čo sa týka indikačnej premennej  $z_t$ , tá sa zvykne definovať ako

$$z_t = \xi_{t-d}, \quad (6)$$

kde  $\xi_t$  predstavuje nejakú pozorovateľnú veličinu a  $d$  je prirodzené číslo.

Výstavba viacrežimového modelu pozostáva podobne ako pri lineárnych modeloch z niekoľkých krokov. Predbežná analýza však už zahrňuje aj testovanie nulovej hypotézy o linearite oproti alternatívnej o 2-režimovej nelinearite. K dispozícii sú dva testy. Prvý navrhol Tsay v [18] na podklade viacrozmerného modelu s premenlivými režimami avšak so skokovým prechodom (TAR model), druhý vznikol na



Obrázok 1: Logistická a exponenciálna prechodová funkcia

jednorozmernom STAR modele a využíva Lagrangeove multiplikátory (LM). Oba počítajú so základnou definíciou indikačnej premennej (6). Testom linearity sa vyberie najlepšia indikačná premenná aj s oneskorením  $d$ . Ďalej nasleduje odhad parametrov a výber najefektívnejšieho modelu, označovaný ako selekcia modelu. Oba úkony možno spojiť do jedného a odhadnúť neznáme parametre minimalizáciou informačných kritérií, ako je napr. Schwarzovo BIC.

## Modelovanie spoločného trendu a sezónnej zložky

Pri modelovaní vektorovej veličiny sa nezriedka stáva, že dve alebo viac jej zložiek obsahuje trend alebo sezónnu zložku. Ak tieto vlastnosti nesú spoločné črty, je možné, že pozorované časové rady generuje ten istý proces. To sa dá využiť na zníženie počtu odhadovaných parametrov či dokonca spresnenie predpovedí. Navyše, vyšetrowanie prítomnosti spoločných vlastností pomáha predchádzať riziku už spomínanej zdanlivej regresie.

Predpokladajme VAR( $p$ ) model  $k$ -rozmerného nestacionárneho časového radu [1], stacionarizovateľný prvými diferenciami,

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Omega}\boldsymbol{\delta}_t + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (7)$$

ktorý obsahuje konštantný člen a deterministickú zložku  $\boldsymbol{\delta}_t$  (s deterministickým trendom alebo sezónnymi premennými), ďalej  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Omega}$  a  $\boldsymbol{\Phi}_i$ ,  $i = 1, \dots, p$ , sú parametre modelu a  $\boldsymbol{\varepsilon}_t$  opäť predstavuje  $k$ -rozmerný gausovský proces bieleho šumu. Tento model po niekoľkých elementárnych úpravách zapíšeme v tvare tzv. modelu korekcie chyby (error-correction, ECM)

$$\Delta\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Omega}\boldsymbol{\delta}_t + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i\Delta\mathbf{y}_{t-i} + \boldsymbol{\Pi}\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad (8)$$

kde  $\boldsymbol{\Gamma}_i = (\boldsymbol{\Phi}_1 + \cdots + \boldsymbol{\Phi}_i) - \mathbf{I}$ , pre  $i = 1, 2, \dots, p-1$ ,  $\boldsymbol{\Pi} = (\boldsymbol{\Phi}_1 + \cdots + \boldsymbol{\Phi}_p) - \mathbf{I}$ , a  $\mathbf{I}$  je jednotková matica. Model ECM obsahuje na jednej strane krátkodobé vzťahy medzi procesmi, čo sú vzťahy medzi diferencovanými (stacionárnymi) procesmi, na druhej strane zahŕňa vzťahy dlhodobé, obsiahnuté v parametrickej matici  $\boldsymbol{\Pi}$ . Tu môžu nastať tri situácie: 1) matica  $\boldsymbol{\Pi}$  má plnú hodnotu, čo znamená, že časový rad je generovaný stacionárnym vektorovým procesom; 2) hodnota  $\boldsymbol{\Pi}$  je nulová, diferencovaním nedochádza ku strate dlhodobých vzťahov medzi časovými radmi,

pretože žiaden neexistuje; 3) hodnosť matice je  $r$ ,  $0 < r < k$ , niektoré časové rady nemôžu byť stacionarizované diferencovaním bez straty informácie, pretože ich lineárne kombinácie s inými časovými radmi už stacionárne sú (tieto rady sa nazývajú kointegrované). V poslednom prípade matica  $\mathbf{\Pi}$  môže byť vyjadrená ako súčin dvoch matíc  $\boldsymbol{\alpha}$  a  $\boldsymbol{\beta}$ , obe s rozmerom  $k \times r$  a hodnosťou  $r$ , platí teda  $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ . Keďže  $\mathbf{y}_t$  je nestacionárny, jeho diferencie  $\Delta\mathbf{y}_t$  sú však stacionárne a lineárna kombinácia  $\boldsymbol{\beta}'\mathbf{y}_t$  taktiež,  $\boldsymbol{\beta}$  možno považovať za kointegračnú maticu zloženú z  $r$  kointegračných vektorov a lineárna kombinácia  $\boldsymbol{\beta}'\mathbf{y}_t$  vyjadruje vzťahy medzi stochastickými procesmi, ktoré sa dlhodobo presadzujú, teda akési ekvilibrium. Mieru odlišnosti dlhodobých a krátkodobých vzťahov udávajú parametre v  $\boldsymbol{\alpha}$  ktoré možno chápať ako silu potrebnú na návrat z krátkodobých výchyliet do ekvilibria. Konštanty  $\boldsymbol{\mu}$  v (8) znamenajú, že  $\mathbf{y}_t$  obsahuje lineárny deterministický trend, a lineárny trend v (8) znamená kvadratický trend v  $\mathbf{y}_t$ .

Na testovanie hodnosti matice  $\mathbf{\Pi}$  sa s výhodou používa viacrozmerná Johansenova metóda [14]. Prepíšme (8) do tvaru pomocnej regresie

$$\mathbf{z}_{0t} = \mathbf{\Gamma}\mathbf{z}_{1t} + \mathbf{\Pi}\mathbf{z}_{2t} + \boldsymbol{\varepsilon}_t. \quad (9)$$

Predpokladajme nateraz, že nestacionaritu  $\mathbf{y}_t$  spôsobuje iba lineárny deterministický trend, potom  $\mathbf{z}_{0t} = \Delta\mathbf{y}_t$ ,  $\mathbf{z}_{1t} = (1, \Delta\mathbf{y}'_{t-1}, \dots, \Delta\mathbf{y}'_{t-p+1})'$ ,  $\mathbf{z}_{2t} = \mathbf{y}_{t-p}$  a  $\mathbf{\Gamma} = (\boldsymbol{\mu}, \mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_{p-1})$ . Zavedením označenia  $\mathbf{M}_{ij} = n^{-1} \sum_t \mathbf{z}_{it}\mathbf{z}'_{jt}$ , pre  $i, j = 0, 1, 2$  a  $\mathbf{S}_{ij} = \mathbf{M}_{ij} - \mathbf{M}_{i1}\mathbf{M}_{11}^{-1}\mathbf{M}_{1j}$ , pre  $i, j = 0, 2$ , odhad kointegračnej matice  $\boldsymbol{\beta}$  dostaneme riešením rovnice

$$|\lambda\mathbf{S}_{22} - \mathbf{S}_{20}\mathbf{S}_{00}^{-1}\mathbf{S}_{02}| = 0, \quad (10)$$

z ktorej sa vypočítajú vlastné čísla  $\hat{\lambda}_1 > \dots > \hat{\lambda}_k$  a k nim prislúchajúce vlastné vektory  $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_k$ . Určenie  $r$  sa vykoná na základe testovacej štatistiky  $LR(r) = -n \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i)$ . Začneme testovaním nulovej hypotézy o žiadnom kointegračnom vzťahu ( $r = 0$ ), ktorú zamietame, ak  $LR(0)$  je menšia ako príslušná tabelovaná kritická hodnota. Postup opakujeme pre zvyšujúce sa  $r$  a končíme pri prvom nezamietnutí. Potom  $\hat{\boldsymbol{\beta}} = (\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_r)$ .

Nájdenie  $r$  kointegračných vzťahov značí, že v  $k$ -rozmernom systéme  $\mathbf{y}_t$  musí existovať  $k - r$  spoločných (vo všeobecnosti stochastických) trendov, ktoré by sme dostali v lineárnej kombinácii  $\boldsymbol{\gamma}'\mathbf{y}_t$ . Odhad  $k \times (k - r)$  matice  $\boldsymbol{\gamma}$  sa vypočíta podobne ako  $\hat{\boldsymbol{\beta}}$ , ale z rovnice  $|\lambda\mathbf{S}_{00} - \mathbf{S}_{02}\mathbf{S}_{22}^{-1}\mathbf{S}_{20}| = 0$ . Jej riešením sú rovnaké vlastné čísla ako z (10), no s inými vlastnými vektormi  $\hat{\mathbf{w}}_1, \dots, \hat{\mathbf{w}}_k$ . Potom  $\hat{\boldsymbol{\gamma}} = (\hat{\mathbf{w}}_{r+1}, \dots, \hat{\mathbf{w}}_k)$  (pozri [11]).

Výšetrovanie spoločnej deterministickej sezónnej zložky v systéme časových radov je založené na veľmi podobnom princípe [5], tentokrát  $\mathbf{z}_{0t} = \Delta\mathbf{y}_t$ ,  $\mathbf{z}_{1t} = (\Delta\mathbf{y}'_{t-1}, \dots, \Delta\mathbf{y}'_{t-p+1}, \mathbf{y}_{t-p})'$ ,  $\mathbf{z}_{2t} = \boldsymbol{\delta}_t$ , kde  $\boldsymbol{\delta}_t$  je  $m$ -rozmerný vektor deterministických sezónnych premenných v čase  $t$ . Hľadáme vlastné čísla súčinu matíc  $\mathbf{S}_{00}^{-1}\mathbf{S}_{02}\mathbf{S}_{22}^{-1}\mathbf{S}_{20}$  zoradené vzostupne,  $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k < 1$ , a im zodpovedajúce vlastné vektory  $\mathbf{u}_1, \dots, \mathbf{u}_k$ . Test redukovanej hodnosti matice  $\boldsymbol{\Omega}$  je znovu testom nulových kanonických korelácií medzi  $\mathbf{z}_{0t}$  a  $\mathbf{z}_{2t}$ . Testovacia štatistika  $LR(r) = -(n-p) \sum_{i=1}^r \ln(1 - \lambda_i)$  má  $\chi^2$  rozdelenie pravdepodobnosti s  $r(k[m - 1] + r)$  stupňami voľnosti za predpokladu platnosti nulovej hypotézy, že  $r$  najmenších korelácií je nulových. Potom  $\mathbf{u}'_i\Delta\mathbf{y}_t$ ,  $i = 1, 2, \dots, r$ , nie sú korelované so žiadnou z lineárnych kombinácií zložiek



sezónnych premenných v  $\delta_t$ , zatiaľ čo  $w'_i \Delta y_t$ ,  $i = r + 1, \dots, k$ , nesú plnú informáciu o sezónnych procesoch.

## Modelovanie závislosti pomocou kopúl

Modelovanie súčasne viacerých náhodných premenných sa nemusí obmedzovať iba na regresiu, kde niektoré premenné sú označené za endogénne a ostatné za exogénne, takisto nás nemusí vždy zaujímať časový vývoj skúmaných javov (ak vôbec nejaký je). V takom prípade skúmame združené rozdelenie systému dvoch alebo viacerých pozorovaných veličín, aby sme dostali informáciu o ich vzájomnom previazaní a tvar marginálneho rozdelenia pravdepodobnosti, ktoré popisuje individuálne vlastnosti každej z nich. Problém nastáva, keď jeho jednorozmerné rozdelenia nie sú z tej istej triedy pre každú náhodnú premennú v systéme, pretože nemôžu tvoriť marginálne klasických združených rozdelení (ako napr. normálneho, gama, či rozdelení extrémnych hodnôt). Modelovanie je takto obmedzené na náhodné premenné s rovnakým jednorozmerným rozdelením pravdepodobnosti, čo je síce v geodézii pri prevahe normálneho rozdelenia meraní bežná, no nie výlučná skúsenosť, keďže interdisciplinarita geodézie počíta s využitím aj napr. hydro-meteorologických pozorovaní.

Modelovanie závislostí založené na kopuliach je ušetrené tohto obmedzenia. Uvažujme pre jednoduchosť dvojicu spojených náhodných premenných  $(X, Y)$  ktorej realizáciu označíme  $(x, y)$ . Podľa známej Sklarovej vety, akákoľvek združená distribučná funkcia  $H(x, y)$  môže byť vyjadrená v tvare

$$H(x, y) = C(F(x), G(y)), \quad x, y \in \mathfrak{R} \quad (11)$$

kde  $F(x)$  a  $G(x)$  sú marginálne distribučné funkcie a  $C : [0, 1]^2 \rightarrow [0, 1]$  je *kopula*. Hlavný prínos je v tom, že takýto prístup umožňuje výber modelu závislosti medzi  $X$  a  $Y$ , reprezentovanom kopulou, nezávisle od voľby marginálnych rozdelení. Model zostane rovnaký i po zmene mierky - akejkolvek monotónnej transformácii - realizácií ľubovolnej z náhodných premenných.

V súčasnosti je teória kopúl pomerne dobre rozpracovaná, rozsiahly úvod možno nájsť napr. v [16], aplikácie sú však stále v počiatkoch, prevažuje oblasť poisťovníctva [9] a hydrológie [10].

## 2 Ciele dizertačnej práce

Ciele a obsah dizertačnej práce možno zhrnúť do nasledujúcich bodov:

- techniky modelovania časových radov, ktoré pomôžu pochopiť osobitosti a/alebo spoločné vlastnosti procesov pôsobiacich v systéme skúmaných veličín;
- viacrozmerné modely vystihujúce premenlivosť režimov v prírodných procesoch;
- modely umožňujúce využiť spoločné zložky v časových radoch na zefektívnenie analýzy a spresnenie predpovedí;

- modely združeného rozdelenia náhodných premenných v pravdepodobnostnom priestore, ktoré osobitne popisujú individuálne vlastnosti premenných a osobitne ich "kolektívne" vlastnosti;
- zlepšenie teoretických možností niektorých modelov, najmä ich zovšeobecnenie;
- poskytnúť širokoučelný a čo najviac úser-friendly" súbor algoritmov pre výpočtový systém Mathematica (postup špecifikácie modelu, test nelinearity, podmienený odhad parametrov viacrežimového modelu, predpovedanie, transformácia podľa spoločných komponentov, modelovanie závislostnej štruktúry v združenom rozdelení).

## 3 Dosaiahnuté výsledky

### 3.1 Viacrežimové modely so spojitým prechodom

Ako sme spomenuli, v TAR a STAR modeloch sa indikačná premenná definuje vzťahom (6) ku ktorému je uvádzaný aj postup výstavby modelu od testovania linearity až po vyhodnotenie adekvátnosti modelu. V našej práci používame všeobecnejší spôsob konštrukcie  $z_t$  a to aplikáciou agregáčného operátora  $\mathcal{A}$ ,

$$z_t = \mathcal{A}(\xi_{t-1}, \dots, \xi_{t-d}). \quad (12)$$

Príkladom agregáčného operátora z  $R^d$  do  $R$  je vážený priemer alebo tzv. usporiadaný vážený priemer (OWA operátor), kde diverzita v definovaní váh pokryje široké potreby praktických aplikácií. Špeciálnym prípadom sú minimum, maximum, aritmetický priemer, ale aj projekcia do súradnice  $\xi_{t-d}$  použitá v (6). Viac o konštrukcii a vlastnostiach agregáčnych operátorov v [3].

Čo sa týka testovania linearity  $\mathbf{y}_t$ , pre Tsayov test sme v našej práci odvodili všeobecnejší postup uvažujúc (12). Začína klasickou lineárnou (auto)regresiou  $\mathbf{y}_t = \Phi \mathbf{X}_t + \varepsilon_t$ , v ktorej jednotlivé vrstvy (usporiadané podľa času) preusporiadame vzostupne podľa narastajúcich hodnôt  $z_t$ , takže  $\mathbf{y}_{t(i)} = \Phi \mathbf{X}_{t(i)} + \varepsilon_{t(i)}$ , kde  $t(i)$  je časový index  $z_{t(i)}$ , teda  $i$ -teho najmenšieho prvku premennej  $z_t$  pre  $i = 1, \dots, n$ , skrátene  $z_{t(i)} = z_{(i)}$ . Takéto usporiadanie neovplyvní dynamiku modelovaných časových radov a navyše transformuje potenciálny dvojrežimový model na jednoduchší problém jediného zlomového bodu, tj. bodu, kde dochádza ku zmene režimov. Ten sa dá detekovať napr. pomocou prediktívnych reziduí  $\varepsilon_{t(i+1)}$ . Za predpokladu linearity  $\mathbf{y}_t$  je rekurzívny odhad z preusporiadanej regresie metódou najmenších štvorcov konzistentný a prediktívne reziduá sa blížia k bielemu šumu, nie sú teda korelované s regresorom  $\mathbf{X}_{t(i+1)}$ . Nech  $\hat{\Phi}_{\tilde{n}}$  je MNS odhad parametra  $\Phi$  z údajov asociovaných s  $\tilde{n}$  najmenšími hodnotami  $z_t$  a

$$\hat{\boldsymbol{\eta}}_{t(\tilde{n}+1)} = \frac{\mathbf{y}_{t(\tilde{n}+1)} - \hat{\Phi}_{\tilde{n}} \mathbf{X}_{t(\tilde{n}+1)}}{[1 + \{\mathbf{X}'_{t(\tilde{n}+1)} \mathbf{V}_{\tilde{n}} \mathbf{X}_{t(\tilde{n}+1)}\}]^{1/2}}$$

sú štandardizované reziduá z 1-krokových predpovedí v preusporiadanej regresii s  $\mathbf{V}_{\tilde{n}} = \left[ \sum_{i=1}^{\tilde{n}} \mathbf{X}_{t(i)} \mathbf{X}'_{t(i)} \right]^{-1}$ . Potom z regresie  $\hat{\boldsymbol{\eta}}_{t(j)} = \Psi \mathbf{X}_{t(j)} + \mathbf{w}_{t(j)}$ ,  $j = \tilde{n}_0 +$

1, \dots, n, kde  $\tilde{n}_0$  označuje štartovací bod rekurzívneho odhadu (štandardne  $\tilde{n}_0 \approx 3\sqrt{n}$ ), testujeme hypotézu  $H_0 : \boldsymbol{\Psi} = \mathbf{0}$  oproti  $H_1 : \boldsymbol{\Psi} \neq \mathbf{0}$  pomocou testovacej štatistiky

$$C = [n - \tilde{n}_0 - K] \times (\ln |\mathbf{S}_0| - \ln |\mathbf{S}_1|), \quad (13)$$

kde

$$\mathbf{S}_0 = \frac{1}{n - \tilde{n}_0} \sum_{j=\tilde{n}_0+1}^n \hat{\boldsymbol{\eta}}_{t(j)} \hat{\boldsymbol{\eta}}'_{t(j)}, \quad \mathbf{S}_1 = \frac{1}{n - \tilde{n}_0} \sum_{j=\tilde{n}_0+1}^n \hat{\boldsymbol{w}}_{t(j)} \hat{\boldsymbol{w}}'_{t(j)}.$$

Za predpokladu platnosti nulovej hypotézy linearít  $\mathbf{y}_t$ ,  $C$  je náhodná premenná s asymptotickým  $\chi^2$  rozdelením pravdepodobnosti s  $kK$  stupňami voľnosti.

Druhý z testov linearít, tzv. LM test sa v dostupnej literatúre prevažne uvádza pre jednorozmerný prípad, podľa našich vedomostí iba [19, s.93–94] poskytuje viac-rozmerné rozšírenie. Ako sme zistili, toto riešenie nesedí pre jeden rozmer, preto sme na základe porovnania s niektorými podobnými štatistickými testami navrhli korekciu v stupňoch voľnosti  $\chi^2$  rozdelenia pravdepodobnosti testovacej štatistiky. Hoci asymptotické vlastnosti takto rozšíreného a upraveného testu neboli teoreticky dokázané, experimentálne skúsenosti na simulovaných údajoch a pomerne dobrá zhoda s Tsayovým testom v detekcii nelinearity nás povzbudzuje k zaradeniu výsledku do tejto kapitoly.

LM test používa kostru viacrežimového modelu so spojitým prechodom na vyjadrenie nulovej hypotézy ako rovnosti  $\boldsymbol{\Phi}_1 = \boldsymbol{\Phi}_2$  alebo  $\gamma = 0$  oproti alternatívnej hypotéze nerovnosti. V súvislosti s testovaním často spomínaný problém neidentifikovaných parametrov v definícii nulovej hypotézy (pozri napr. [8]) tu je vyriešený aproximáciou prechodovej funkcie  $G(z_t, \gamma, r)$  Taylorovým polynómom vhodného rádu okolo  $\gamma = 0$ . V prípade logistického STAR najvhodnejším sa javí tretí rád ([15]), zatiaľčo druhý rád sa odporúča pre exponenciálnu prechodovú funkciu ([6]). Po úpravách dostaneme pomocnú regresiu

$$\mathbf{y}_t = \boldsymbol{\beta}_{0,0} + \boldsymbol{\beta}'_0 \mathbf{X}_t + \boldsymbol{\beta}'_1 \mathbf{X}_t z_t + \boldsymbol{\beta}'_2 \mathbf{X}_t z_t^2 + \boldsymbol{\beta}'_3 \mathbf{X}_t z_t^3 + \mathbf{e}_t, \quad (14)$$

pre logistický prechod a (14) s extra členom  $\boldsymbol{\beta}'_4 \mathbf{X}_t z_t^4$  pre exponenciálny. Parametre  $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, \dots, \beta_{i,K-1})'$ ,  $i = 0, 1, 2, 3, 4$ , sú funkciami parametrov  $\boldsymbol{\Phi}_1$ ,  $\boldsymbol{\Phi}_2$ ,  $\gamma$  a  $r$ . Nulová hypotéza  $\gamma = 0$  teraz zodpovedá  $H'_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 (= \boldsymbol{\beta}_4) = \mathbf{0}$  (tiež  $\mathbf{e}_t = \boldsymbol{\varepsilon}_t$ ). Všimnime si, že ak  $z_t$  je jednou z premenných zahrnutých v  $\mathbf{X}_t$ , členy  $\beta_{i,0} z_t^i$ ,  $i = 1, 2, 3, 4$ , musia byť z pomocnej regresie vylúčené, aby sa predišlo úplnej multikolinearite.

Testovacia štatistika sa vypočíta ako

$$LM = n(\ln |\boldsymbol{\Sigma}_0| - \ln |\boldsymbol{\Sigma}_1|) \quad (15)$$

kde

$$\boldsymbol{\Sigma}_0 = n^{-1} \sum_t \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t, \quad \boldsymbol{\Sigma}_1 = n^{-1} \sum_t \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}'_t,$$

pričom  $\hat{\boldsymbol{\varepsilon}}_t$  sú reziduá odhadnuté z lineárneho modelu  $\mathbf{y}_t = \boldsymbol{\Phi} \mathbf{X}_t + \boldsymbol{\varepsilon}_t$  a  $\hat{\boldsymbol{\varepsilon}}_t$  z regresie  $\hat{\boldsymbol{\varepsilon}}_t$  od  $\mathbf{X}_t$  a  $\mathbf{X}_t z_t^i$ ,  $i = 0, 1, 2, 3, (4)$ . Za predpokladu platnosti nulovej hypotézy,  $LM$  má asymptoticky  $\chi^2$  rozdelenie s  $3kK$  stupňami voľnosti v prípade logistického prechodu, a  $4kK$ , ak uvažujeme exponenciálny STAR ako alternatívny model.

Jedným z krokov výstavby modelu po testovaní linearity je výber modelu s pomocou informačného kritéria, napríklad Schwarzovho BIC. Verzia BIC, ktorou sme určili rád lineárneho VAR modelu pre účely testovania linearity, už nie je dostatočujúca pre určenie rádu autoregresie STAR modelov, a to ani keby boli rády či dokonca počty observácií v jednotlivých režimoch rovnaké. V [18] sa uvádza informačné kritérium, ktoré rešpektuje rozdielnosti v (dvoch) režimoch, a to už aj viacrozmerného, no stále iba TAR modelu. V predkladanej práci zovšeobecňujeme výpočet BIC na platnosť pre  $k$ -rozmerný  $m$ -režimový autoregresný model so spojitým prechodom.

Rozšírime model (4) do  $m$  režimov (iné možnosti sú napr. v [17]), každý s vlastným regresorom, teda

$$\mathbf{y}_t = \Phi_1 \mathbf{X}_{1t} (1 - G(z_t, \gamma_1, r_1)) + \Phi_2 \mathbf{X}_{2t} (G(z_t, \gamma_2, r_2) - G(z_t, \gamma_3, r_3)) + \dots + \Phi_m \mathbf{X}_{mt} (G(z_t, \gamma_{m-1}, r_{m-1}) - G(z_t, \gamma_m, r_m)) + \varepsilon_t, \quad (16)$$

a definujeme regresor v  $j$ -tom režime,

$$\mathbf{X}_{j,t} = (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p_j}, \mathbf{x}'_{t-1}, \dots, \mathbf{x}'_{t-q_j})', \quad j = 1, 2, \dots, m, \quad (17)$$

kde  $\mathbf{x}_t$  symbolizuje  $l$ -rozmernú exogénnu premennú,  $p_j$  je rád endogénnej a  $q_j$  exogénnej regresie, potom  $K_j = kp_j + lq_j + 1$  je rozmer regresoru  $\mathbf{X}_{j,t}$ . Počet  $n_j$  observácií v konkrétnom režime nemusí byť nutne prirodzené číslo, ide skôr o váhu ( $\sum_{j=1}^m n_j = n$ ) definovanú ako  $n_j = \sum_t \Delta G_{j,t}$  s  $\Delta G_{j,t} = G_{j-1,t} - G_{j,t}$ , kde  $G_{j,t} = G_j(z_t, \gamma_j, r_j)$  je prechodová funkcia zodpovedajúca  $j$ -temu režimu,  $G_{0,t} = 1$  a  $G_{m,t} = 0$ . Potom Schwarzovo informačné kritérium selekcie  $k$ -rozmerného  $m$ -režimového STAR modelu

$$\mathbf{y}_t = \sum_{j=1}^m \Phi_j \mathbf{X}_{j,t} \Delta G_{j,t} + \varepsilon_t \quad (18)$$

možno zapísať ako

$$BIC(\mathbf{p}, \mathbf{q}) = \sum_{j=1}^m \left( n_j \ln |\hat{\Sigma}_j| + \ln(n_j) k K_j \right), \quad (19)$$

kde  $\mathbf{p} = (p_1, \dots, p_m)$ ,  $\mathbf{q} = (q_1, \dots, q_m)$ , a

$$\hat{\Sigma}_j = \frac{1}{n_j} \sum_t (\mathbf{y}_t - \hat{\Phi}_j \mathbf{X}_{j,t})(\mathbf{y}_t - \hat{\Phi}_j \mathbf{X}_{j,t})' \Delta G_{j,t}$$

je odhad kovariančnej matice reziduí. Ďalšie informačné kritériá, líšiac sa iba penalizačnou funkciou (napr. AIC alebo HQIC), sa vypočítajú podobným spôsobom.

Pre úplnosť dodajme, že odhad parametrov  $\boldsymbol{\theta} = (\Phi, \boldsymbol{\gamma}, \mathbf{r})$ , kde  $\Phi = (\Phi_1, \dots, \Phi_m)$ ,  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{m-1})$ ,  $\mathbf{r} = (r_1, \dots, r_{m-1})$  sa vykoná minimalizáciou stopy kovariančnej matice  $\Sigma(\boldsymbol{\theta}) = \sum_t (\mathbf{y}_t - \Phi \mathbf{X}_t)(\mathbf{y}_t - \Phi \mathbf{X}_t)'$ , kde  $\mathbf{X}_t(\boldsymbol{\gamma}, \mathbf{r}) = (\mathbf{X}'_{1,t} \Delta G_1, \dots, \mathbf{X}'_{m,t} \Delta G_m)'$  a  $\Delta G_j$  je opäť funkciou parametrov  $(\gamma_j, r_j)$ . Jeden spôsob je použiť priamo nelineárnu metódu najmenších štvorcov

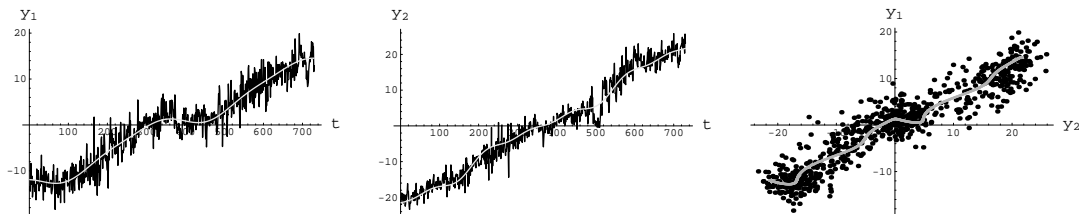
$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \operatorname{Tr}(\Sigma(\Phi, \boldsymbol{\gamma}, \mathbf{r})),$$

ktorá je však najmä vďaka parametru  $\gamma$  výpočtovo náročná a nestabilná, alebo použiť metódu podmienených odhadov, kde sa klasickou lineárnou MNS odhadujú parametre  $\Phi$  závisle na pevných hodnotách  $\gamma$  a  $\mathbf{r}$ .

Alternatívou k výberu modelu podľa stopy, resp. determinantu kovariančnej matice reziduí môže byť porovnanie strednej predpovednej chyby pomocou modifikovaného Diebold-Marianovho testu podľa [13], v ktorom tzv. "loss differential" vo viacrozmernom prípade navrhujeme v tvare L2-normy vektora chýb predpovedí v adekvátnej časovej vrstve.

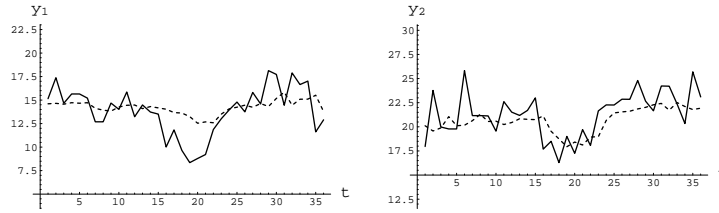
Horeuvedené postupy boli použité aj v nami vytvorenom súbore programových nástrojov pre vybudovanie STAR modelov, ako metóda odhadu parametrov je použitá druhá zo spomínaných, teda podmienená. Popis programu je uvedený na konci kapitoly.

### 3.2 Spoločný trend a sezónna zložka



Obrázok 2: Časové rady GPS pozorování polohy bodu na permanentnej stanici BOR1 s lineárnym trendom a sezónnou zložkou, v jednotkách [mm/deň].

Vo vlastnom experimente sme použili dvojrozmerný časový rad observácií horizontálnej polohy bodu a na základe presnosti modelovania 1-krokových predpovedí porovnávali 3 modelovacie metódy: 1) bežne používaný jednorozmerný AR model, 2) dvojrozmerný VAR a napokon 3) jednorozmerný AR aplikovaný na časové rady, ktoré vznikli dekompozíciou  $\mathbf{y}_t$  na zložku obsahujúcu spoločné deterministické vlastnosti pôvodných radov a zložku, ktorá ich neobsahuje. Dekompozícia je v podstate transformácia z jedného súradnicového systému do druhého, v prípade trendu dostaneme súradnice v novom systéme z  $\mathbf{y}_t^T = (\mathbf{w}_2, \mathbf{v}_1)' \mathbf{y}_t$ , prvá zložka nesie dlhodobú pamäť a druhá predstavuje ekvilírium. V prípade deterministickej sezónnosti dostávame transformáciu  $\mathbf{y}_t^S = (\mathbf{u}_2, \mathbf{u}_1)' \mathbf{y}_t$ , z ktorej podobne  $y_{1t}^S$  nesie plnú informáciu o sezónnom prejave generujúceho procesu,  $y_{2t}^S$  žiadnu. Namodelované predpovede boli transformované naspäť do systému pôvodného časového radu a z takýchto vypočítaná stredná predpovedná chyba MSPE. Postup sme odskúšali na dvojročnom časovom rade s denným rozstupom a sezónnou zložkou modelovanou trigonometrickým polynómom (Obr.2), aj na 43 dní dlhom časovom rade trojhodinových GPS riešení s ôsmimi umelými sezónnymi premennými. V oboch, pri zachovaní rovnakého počtu parametrov vo všetkých porovnávaných lineárnych modeloch, vyšli stredné chyby predpovedí transformovaných časových radov výrazne lepšie (o 5% až 11% menšie) ako MSPE pôvodných, štatistický rozdiel potvrdil aj Diebold-Marianov test. Na Obr.3 sú vykreslené predpovede popri skutočných hodnotách časového radu.



Obrázok 3: Skutočné hodnoty časového radu a predpovede (čiarkovane) v čase  $694 + t$ , [mm/deň].

### 3.3 Modelovanie závislosti pomocou kopúl

V predkladanej práci sa zameriavame aj na objasnenie základnej idey kopúl a fundamentálnych techník odhadu jednoparametrickej triedy Archimedovských kopúl, ktoré sú vďaka jednoduchej konštrukcii, niekoľkým dobrým vlastnostiam a rozmanitosti široko používané. Klasickú neparametrickú a semiparametrickú metódu odhadu sledovaného parametra kopule porovnávame so spôsobom odhadu pomocou nelineárnej regresie a empirickej kopule, ktorý je univerzálnejší, i keď výpočtovo podstatne náročnejší.

Okrem toho v práci navrhujeme spresnenie modelu závislostí cez konvexnú kombináciu dvoch kopúl  $C_1, C_2$ ,

$$C_\alpha(u, v) = \alpha C_1(u, v) + (1 - \alpha) C_2(u, v), \quad (20)$$

kde  $u = F(x)$  a  $v = G(x)$ . Ide o jednoduchú a prakticky dobre interpretovateľnú konštrukciu parametrickej triedy kopúl. Vyšetrením vlastností možno ľahko ukázať, že  $C_\alpha$  je naozaj kopula. Výrazné zlepšenie v popísaní reálnych dát takouto kopulou oproti jednoparametrickým dokladujeme porovnaním súčtu štvorcov reziduí z fitovania na empirickú kopulu.

### 3.4 Výstavba modelu s programom v systéme Mathematica

Súčasťou dizertačnej práce je aj súbor výpočtov v systéme Mathematica rozdelený do štyroch pracovných hárkov ("notebookov"):

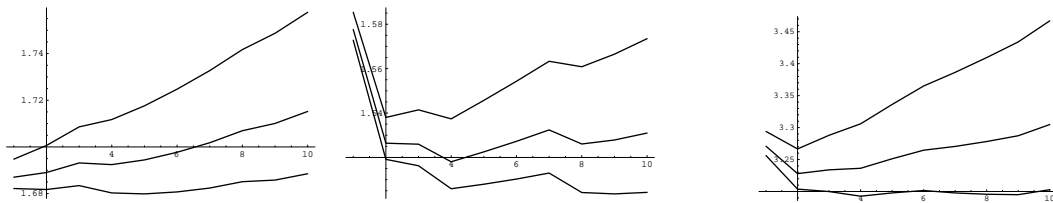
1. Výstavba dvojrežimových modelov s aplikáciou najprv na umelo generované dvojrozmerné TAR-nelineárne stacionárne časové rady a podrobným popisom predbežnej analýzy, testovania časových radov, špecifikácie a vyhodnotenia modelu. Následná aplikácia na reálne časové rady GPS pozorovaní je podaná v ucelenom súhrne.
2. Lineárne modelovanie dvojrozmerných nestacionárnych časových radov obsahujúcich deterministický trend a deterministickú sezónnu zložku. Na základe presnosti predpovede sú porovnávané tri základné prístupy: (a) dvojrozmerný VAR model, jednorozmerný AR model aplikovaný na (b) pôvodné a (c) transformované časové rady so segregovanými deterministickými zložkami.
3. Definície (zdrojový kód) všetkých výpočtových funkcií použitých v predošlých dvoch hárkoch. V takejto podobe sa funkcie musia pred použitím vyhodnotiť

manuálne po načítaní potrebných balíkov systému Mathematica. Jednoduchšia alternatíva je spustiť zapuzdrenú ("balíkovú") formu, ktorá sa dá stiahnuť z autorovej webstránky [www.math.sk/bacigal/homepage/research\\_sk.html](http://www.math.sk/bacigal/homepage/research_sk.html).

4. Analýza závislostnej štruktúry vektorovej náhodnej premennej. V troch sekciách je uvedený fitovací algoritmus troch jednoparametrických Archimedovských kopúl na reálne dáta, ďalej uvažujeme lineárne rozšírenie na trojparametrickú kopulu a nakoniec je uvedený výpočet štandardných odchyliiek odhadu parametrov z neparametrickej a semi-parametrickej metódy.

V nasledujúcich odstavcoch popíšeme potenciál využitia programových rutín v konkrétnej fáze modelovania časových radov. Detailnejšie informácie sa dajú získať priamo z poznámok v notebookoch. Poznamenajme ešte, že pomenovanie nami definovaných funkcií sa riadi podobným princípom ako názvy vstavaných funkcií s tým rozdielom, že začína malým písmenom *f*.

### Predbežná analýza, lineárne modelovanie



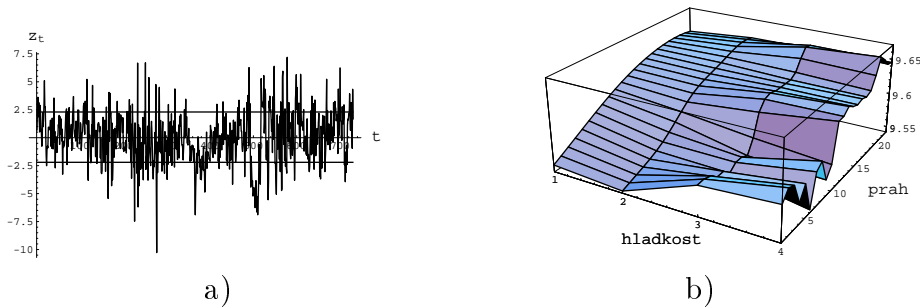
Obrázok 4: Zhora BIC, HQIC a AIC pre výber rádu AR modelov a VAR modelu

Notebooky sa vo všeobecnosti začínajú úvodnými nastaveniami, načítaním balíkov a súborov údajov, nasleduje predbežná analýza vykreslením sériových korelácií a spektrálnych hustôt. Pre nájdenie rádu lineárnych AR alebo VAR modelov sa pomocou Levinson-Durbinovej procedúry (balík `TimeSeries`) vypočítajú AIC, BIC a HQIC (Obr.4). Alternatívne poskytujeme rozšírenie tejto procedúry aj o počítanie s exogénnymi časovými radmi. Ak je časový rad nestacionárny v dôsledku deterministických zložiek, menovite polynomický trend (konštantný, lineárny, kvadratický...) alebo sínusové vlny, a splnenie podmienky stacionarity je nevyhnutné pre ďalšie spracovanie, funkcia `fDeterministicsRemoval` urobí dekompozíciu a vráti reziduá. Regresia nepravidelnej sezónnej zložky, cyklickej zložky a autoregresia sa dá urobiť cez regresor a `fΦ`.

### Premennivé režimy

K dispozícii sú dva testy linearity, `fTsayLTest` pre testovanie oproti alternatíve skokového prechodu a `fLMtypeLTest` oproti alternatívne modelu so spojitým prechodom (buď logistickým alebo exponenciálnym). Obe funkcie umožňujú zapojenie agregáčnych operátorov (niektoré sú preddefinované v `fAggregationOperator`) aj exogénnych premenných, výstupom je testovacia štatistika a stupne voľnosti  $\chi^2$  rozdelenia, ktoré sú pomocou `fChiSquarePValue` prevedené na p-hodnotu. Ak testujeme cez mriežku vstupných parametrov (rád autoregresie  $p, q$  a maximálne oneskorenie  $d$ ), `fArgMinTensor` nájde najmenšiu p-hodnotu a príslušné parametre.

Odhad modelu s premenlivými režimami je úlohou `fConditionalRegimeSwitching`. Ako názov prezrádza, je použitý klasický lineárny odhad MNŠ (podmieneny pevnými hodnotami oneskorenia  $d$ , prahovej hodnoty  $r$  a parametra hladkosti prechodu  $\gamma$ ) namiesto nelineárneho odhadu všetkých neznámych parametrov v jednom kroku. Funkcia má sedem argumentov, umožňuje vstup endogénnych i exogénnych premenných do regresie, tiež ich minulých hodnôt, ktorých časový posun možno ľubovoľne zadať, či už pomocou maximálnej hodnoty (rádu) alebo vymenovaním, pre každú premennú zvlášť i pre všetky spolu, pre všetky režimy naraz i pre každý zvlášť. Parametre  $(d, r, \gamma)$  tvoria piaty argument, buď ako zoznam diskrétnych hodnôt, alebo rovnomerná postupnosť daná hranicami a krokom. Tu sa zide `fThresholdRange`, ktorá z centrálneho pásu indikačnej premennej vyberie postupnosť prahových hodnôt (Obr.5a). Ak uprednostňujeme skokový, nespojitý prechod medzi režimami, stačí vynechať hodnoty  $\gamma$  zo zoznamu parametrov. Posledný argument určuje celkové zameranie funkcie a môže obsahovať ďalšie vstupné údaje. Jeho prvý element slúži ako prepínač medzi módm (vysvetlenie nižšie), medzi logistickou a exponenciálnou prechodovou funkciou, a medzi voľbami konštrukcie indikačnej premennej. Druhý element nesie meno agregáčného operátora a tretí pokrýva potreby jednotlivých módo. Mód funkcie `fConditionalRegimeSwitching` nie je nič iné ako konkrétna úloha, ktorú má vykonať, a výstup, ktorý má vrátiť. V súčasnej verzii je dostupných 5 módo (výstupov): 1) suma štvorcov reziduí ako stopa kovariančnej matice, 2) odhad matice neznámych parametrov  $\hat{\Phi}$ , reziduí  $\hat{\varepsilon}_t$  a ich kovariančnej matice  $\hat{\Sigma}$ , 3) AIC a BIC kritériá výberu, 4) skutočné chyby predpovedí a nakoniec 5) predpovede. Chyby predpovedí sú počítané na vzorke useknutej z konca modelovaného časového radu, jej dĺžka musí byť zadaná spolu s horizontom predpovedí a počtom Monte Carlo cyklov. Ak chceme predpovede z regresie, do ktorej vstupujú exogénne premenné, tieto musia byť dopredu vygenerované a uvedené vo vstupoch funkcie.



Obrázok 5: a) Rozsah prahových hodnôt. b) Výsledok módu 1 podmieneného odhadu (suma štvorcov reziduí, vrstva  $d = 1$ ).

Výsledky sú štrukturované ako vnorené listy podľa poradia parametrov  $d, r_1, \dots, r_{m-1}, \gamma_1, \dots, \gamma_{m-1}$  (v prípade  $m$  režimov), hodnoty parametrov sú umiestnené v druhej časti výstupu. Pre módy 1) a 3) funkcia `fArgMinTensor` prehľadáva prvú časť výstupu a vráti parametre prislúchajúce globálnemu minimu. Tie potom vstupujú do módu 2) a výsledok je podrobený testu zvyškových korelácií (`fPortmanteauTest`) a linearity (`fResidualTest`). Umiestnenie minimálnej hodnoty vo výsledku prvého a tretieho módu je dobré vyšetriť aj vizuálne (Obr.5b) pre prípad, že nájdené globálne minimum nezodpovedá interpretácii modelu alebo jeho hodnota nie je dostatočne



signifikantná.

## Porovnanie presnosti predpovedí

Schopnosť modelov postihnúť priebeh veličiny aj mimo modelovaný úsek môžeme vyhodnotiť napr. zostavením "hitparády". Pomocou `fMeanXPredictionError-Hitparade` je pozícia jednotlivých modelov priamo určená ich dosiahnutou MSPE, MAPE alebo inou mierou úspešnosti fitovania. Naproti tomu `fDieboldMariano-Hitparade` najprv štatisticky testuje rovnosť predpovednej presnosti dvojíc modelov, zostaví tabuľku priebežného skóre a napokon vráti poradie, v ktorom môže jednu pozíciu obsadiť i viac ako jeden "súťažiaci". Vstupom pre obe funkcie sú chyby predpovedí a tzv. "loss" funkcia, ktorá spracováva chyby predpovedí v jednotlivých (časových) vrstvách predtým, než sa zosumujú a podelia počtom. Diebold-Marianov test navyše vyžaduje vstup horizontu predpovedí a hladinu významnosti.

## Odhad parametrov kopule

Programové rutiny v poslednom notebooku nepotrebujú veľmi komentár, nakoľko pozostávajú z jednoduchých príkazov a vstavaných funkcií. Zaujímavým môže byť porovnanie času potrebného na zbehnutie jednotlivých metód, poskytovaného funkciou `Timing`. Najzdlhavesšia procedúra je bezkonkurenčne generovanie empirickej kopule s celkovým časom okolo pol hodiny na 1GHz počítači. Toto je zjavne jeden z hlavných nedostatkov systému Mathematica, ktorý v efektívnosti narábania s veľkým množstvom údajov nedokáže konkurovať programom priamo písaným napr. v jazyku C. Sila "počítačového algebrového systému" Mathematica stojí skôr vo výbornom pomere univerzálnosti využitia ku jednoduchosti programovania, zvlášť na poli symbolického návrhu experimentov.

O program na fitovanie kopúl prejavil záujem aj zahraničný kolega z University of Delaware, procedúry na výstavbu modelov s premenlivými režimami v súčasnosti už používa niekoľko študentov tretieho stupňa vysokoškolského štúdia na Stavebnej fakulte STU a Prírodovedeckej Fakulte UK v Bratislave v ich aplikovanom výskume, a študent magisterského štúdia FMFI UK ku záverečnej diplomovej práci.

## Perspektívy ďalšieho výskumu

Analýza časových radov je neustále sa rozvíjajúca disciplína štatistiky, mnohé z problémov načrtnutých v dizertačnej práci čakajú na seriózne matematické podloženie, iné si žiadajú praktický dôkaz ich využiteľnosti v aplikáciách technických a spoločenských vied. V našom ďalšom výskume vidíme priestor pre dosiahnutie nasledovných cieľov:

- implementovať ďalšie možnosti rozšírenia modelov s premenlivými režimami do troch a viac režimov a s aplikovaním nových tried agregáčnych operátorov;
- modelovať spoločnú nelinearitu a skúmať prítomnosť spoločných prvkov v režimoch;

- porovnať efektívnosť techník na fitovanie viacparametrických kopúl, preskúmať praktické vlastnosti tzv. Archimax kopúl modelujúcich nesymetrické závislosti náhodných premenných;
- nájsť rozumné aplikácie v geodézii a príbuzných vedách v zmysle interpretácie dosiahnutých výsledkov;
- sprístupniť program aj pre iných ako používateľov komerčného systému Mathematica, konkrétne prevedením kódu do podoby webaplikácie alebo niektorého z voľne šíriteľných programovacích platforiem (napr. *R* alebo jazyk *C*)

## Zoznam publikácií a citácií

- V1 Bacigál, T.: Multivariate Threshold Autoregressive Models in Geodesy, *Journal of Electrical Engineering* 12/S, vol. 55, 91–94 (2004).
- Komorník, J., Komorníková, M.: Application of aggregation operators in regime-switching models for multiple exchange rates, *Proc. IPMU 2006*
  - Komorník, J., Komorníková, M.: Applications of regime-switching models based on aggregation operators, prijaté do *Kybernetika*
- V2 Bacigál, T.: Modeling point's position time series with respect to common trend, *Zborník príspevkov konferencie PRASTAN*, Bratislava, 5–10 (2004).
- V3 Bacigál, T., Komorníková, M.: Modelling point's position time series in the light of cointegration, *Proceedings of INGENO 2004 (CD-ROM edition)*, Bratislava (2004).
- V4 Bacigál, T., et al.: Variance components estimate in 2D geodetic network, *Proceedings of INGENO 2004 (CD-ROM edition)*, Bratislava (2004).
- V5 Bacigál, T.: Testing for Common Deterministic Trends in Geodetic Data, *Journal of Electrical Engineering* 12/S, vol. 56, 115–118 (2005).
- V6 Bacigál, T.: Modelling Relationship Using Archimedean Copula: An Introduction to Experimental Study, *Proceedings of MAGIA 2005*, Bratislava, 60–72 (2005).
- V7 Bacigál, T.: Fitting Archimedean copulas to bivariate geodetic observations, *Proceedings of APLIMAT 2006*, Bratislava, 519–527 (2006).
- V8 Bacigál, T., Komorníková, M.: Fitting Archimedean copulas to bivariate geodetic data, *Proceedings in Computational Statistics COMPSTAT 2006*, Rome, Physica-Verlag Heidelberg, Springer, 649–656 (2006).
- V9 Bacigál, T.: Multivariate LSTAR in geodesy, *Journal of Electrical Engineering* 12/S, vol. 57, 115–118 (2006).
- V10 Bacigál, T.: Multivariate smooth transition AR model with aggregation operators and application to exchange rates, accepted to *Kybernetika* (2007).

## Literatúra

- [1] Arlt, J.: Moderní metody modelování ekonomických časových řad, Grada publishing, Praha (1999).
- [2] Arlt, J., Arltová, M.: Finanční časové řady, Grada publishing, Praha (2003).
- [3] Calvo, T., Kolesárová, A., Komorníková, M., Mesiar, R.: Aggregation operators: Properties, classes and construction methods. In: Aggregation operators (T. Calvo, G. Mayor and R. Mesiar, eds.), Physica-Verlag, New York, pp.3–140 (2002).
- [4] Cipra, T.: Analýza časových řad s aplikacemi v ekonomii, Alfa, Praha (1986).
- [5] Engle, R.F., Hylleberg, S.: Common Seasonal Features: Global Unemployment, Oxford Bulletin of Economics and Statistics, vol. 58, 615–630 (1996)
- [6] Escribano, A., Jordá, O.: Improved Testing and specification of smooth transition regression models. In: Nonlinear time series analysis of economic and financial data (P. Rothman, ed.), Kluwer, Boston 1999, pp.289–319.
- [7] Franses, P.H.: Time Series Models for Business and Economic Forecasting, Cambridge University Press, Cambridge (1998).
- [8] Franses, P.H., van Dijk, D.: Non-linear Time Series Models in Empirical Finance, Cambridge University Press, Cambridge (2000).
- [9] Frees, E.W., Valdez, E.A.: Understanding Relationships Using Copulas, North American Actuarial Journal, vol. 2, 1–25 (1998).
- [10] Genest, C., Favre, A.-C.: Everything you allways wanted to know about copula modeling but were afraid to ask, ??? (2006).
- [11] Gonzalo, J., Granger, C.W.J.: Estimation of Common Long-Memory Components in Cointegrated Systems, Journal of Business and Economic Statistics, vol. 13, 27–36 (1995).
- [12] Granger, C.W.J., Teräsvirta, T.: Modelling Nonlinear Economic Relationships, Oxford University Press, Oxford (1993).
- [13] Harvey, D.I., Leybourne, S.J., Newbold, P.: Testing the Equality of Prediction Mean Squared Errors, International Journal of Forecasting, vol. 13, 281–291 (1997).
- [14] Johansen, S.: Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, Econometrica, vol. 59, 1551–1580 (1991).
- [15] Luukkonen, R., Saikkonen, P., Teräsvirta, T.: Testing linearity against smooth transition autoregressive models, Biometrika, vol. 75, 491–499 (1988).

- [16] Nelsen, R.B.: An Introduction to Copulas, Lecture Notes in Statistics, vol. 139, Springer (1998).
- [17] van Dijk, D., Teräsvirta, T., Franses, P.H.: Smooth Transition Autoregressive Models - A Survey of Recent Developments, Econometric Institute Research Report EI 2000-23/A (2000).
- [18] Tsay, R.S.: Testing and modeling multivariate threshold models, Journal of the American Statistical Association, vol. 93, 1188–1202 (1998).
- [19] Weise, C.L.: The Asymmetric Effects of Monetary Policy: A Nonlinear Vector Autoregression Approach, Journal of Money, Credit, and Banking, vol. 31, 85–108 (1999).

## Summary

Multivariate modelling methods count doubtlessly among the modern approaches in time series processing, allowing us to capture wider area of causalities that are of our interest. Supported by powerful data collecting technologies such as GPS, we are able to model relations of time space entities, namely the varying geometrical positions of points. They reflect processes that interact locally, regionally and globally. In the thesis, we have presented several known methods and proposed some new methods for multivariate time series modelling, as listed below.

In terms of particular application, we concentrated on permanent GPS observations, that serve for regular monitoring of the Earth's crust kinematics (and for other research purposes). The time series of north and east horizontal coordinates clearly show the long-term drift of Eurasian tectonic plate as linear deterministic trend, which is common to every point in the area. Besides this, the time series visibly reflect other effects (e.g., deterministic seasonality), either as a common or as a unique feature. They can be and partly they were distinguished from each other by the methods of multivariate modelling as proposed in this thesis.

The work on the thesis brought several contributions both to theoretical and practical aspects of time series analysis.

First, the thesis has good informative value due to summarizing the recent methods in multivariate context.

Second, we have proposed new extension to regime-switching models when transition variable is determined by aggregation operators. Thanks to variety of such operators, usefulness of this approach is appreciable as it gives more freedom to fit one's empirical feeling without any loss of reliability. The extension causes no limitation to properties of the nonlinear models so the familiar testing and parameter estimation procedures apply as shown in a standalone section.

Third, a more general formula has been derived for Akaike and Schwarz information criteria used to choose the correct order of VAR in individual regimes. The generalisation allows computer aided modelling to encompass vector variable, arbitrary transition function and regime varying order of autoregression in single procedure.

Fourth, we extend the (modified) Diebold-Mariano test to multivariate case as - to our best knowledge - no such attempt has been published so far.

Fifth, in experimental part of thesis we have shown that investigating common features can significantly improve forecast accuracy of linear models. Improvement in fit is achieved also in using linear convex combination of two copulas, which can be considered as very easy and effective way of constructing multi-parametric copulas. This approach is brand new to copula theory.

At last, the complete source code is available in appendix for the main modelling routines, that is nonlinear model specification, cointegration and common seasonality detection and transformation, nonparametric and semi-parametric fitting of Archimedean copula compared to more universal nonlinear least squares based fit approach. All the routines are coded in computer algebra system Mathematica, properly supplied with description to help practitioners easily implement their specific ideas. The copula fitting program has already been requested from abroad for app-

lication in civil engineering while routines for modelling nonlinearity are nowadays used by several postgradual students from Slovakia in their research.

Among all mentioned modelling procedures, our prime attention was paid to the largest algorithm encapsulated to function `fConditionalRegimeSwitching`. It focuses on conditional estimation of general STAR model parameters and allows one to set variety of input options according to particular problem, e.g., inclusion of exogenous variables, regime specific AR orders, number of regimes, parameters grid, transition and aggregation functions, number of forecast steps and Monte-Carlo cycles. As output - besides parameters estimation and forecasting - it offers the measures for model selection based on both in-sample (sum of squared residuals, information criteria) and out-of-sample fit (forecast errors).

Basically, the whole work is tried to be elaborated keeping in mind the best possible universality thus allowing wide applicability in practice and convenient implementation into computer languages, for instance either as web-based application accessible on Internet or locally executable programme built from *C* source code export.