METODE ODKRIVANJA GROBIH POGREŠKOV V GEODETSKIH OPAZOVANJIH

METHODS OF GROSS ERROR DETECTION IN GEODETIC OBSERVATIONS

Dejan Grigillo, Bojan Stopar

UDK: 528.1

Klasifikacija prispevka po COBISS-u: 1.02

POVZETEK

V članku opisujemo tri metode ugotavljanja in iskanja morebiti grobo pogrešenih opazovanj. Osnovo za vse metode predstavljajo nadštevilna opazovanja. Globalni test modela s t. i. Data Snoopingom je najpogosteje uporabljena metoda za iskanje grobo pogrešenih opazovanj, vendar zahteva zanesljivo poznavanje stohastičnih lastnosti opazovanj. Zato sta kot alternativi predstavljena še t. i. test τ in t. i. danska metoda. Na enostavnem primeru ravninskega geodetskega četverokotnika, z opazovanimi šestimi dolžinami in tremi horizontalnimi koti, so predstavljene vse tri metode iskanja grobih pogreškov.

ABSTRACT

The article describes three methods of gross error detection and their localization in geodetic surveying. The prerequisite for any gross error detection procedure is the availability of a set of redundant observations. The global model test with Data Snooping is the most commonly used method for gross error detection, however, it assumes that the a priori precision of observations is reliably known. As alternatives, the τ test and the Danish method are presented. An example of gross error detection in a plane cross-braced quadrilateral is given for all three methods.

KLJUČNE BESEDE

grobi pogrešek, izravnava po metodi najmanjših kvadratov, statistična hipoteza, Data Snooping, test τ, danska metoda

KEY WORDS

gross error, least squares adjustment method, statistical hypothesis, Data Snooping, τ test, Danish method

UVOD

Tehnološki razvoj prinaša vedno kakovostnejši in vse bolj prefinjen merski instrumentarij, merski postopki in obdelava opazovanj postajajo vse bolj avtomatizirani. To po eni strani geodetu omogoča kakovostno, hitro in dokaj enostavno izvedbo izmere ter vrednotenje pridobljenih rezultatov, po drugi strani pa postajata merski instrumentarij in obdelava podatkov opazovanj vedno bolj domena »črne skrinjice«. Poleg prednosti se tako tudi zmanjšuje operaterjev nadzor nad izvedbo meritev, pridobljenimi rezultati in njihovo kakovostjo. Operater mora v taki situaciji pogosto preprosto zaupati pridobljenim rezultatom.

Kljub tehnološkemu napredku pa terenska opazovanja, zapis in obdelava opazovanj niso in nikoli ne bodo idealni. Merski postopek obremenjujejo številni vplivi, rezultat katerih so pogreški opazovanj. Klasična teorija pogreškov loči slučajne, sistematične in grobe pogreške. Slučajni pogreški so neizbežni in jih obravnavamo kot naravno lastnost opazovanj. Sistematične pogreške lahko z uporabo kalibriranega ali kompariranega instrumentarija, z ustrezno metodo izmere ali z modeliranjem v matematičnem modelu izravnave v celoti odpravimo. Grobi pogreški so rezultat malomarnosti operaterja in/ali nepravilnosti delovanja merskega instrumenta.

Odkrivanje grobih pogreškov, nastalih med zapisom opazovanj, prenosom podatkov, njihovo obdelavo in izravnavo, je lahko drago in dolgotrajno, če se ga ne lotimo sistematično. Da bodo rezultati geodetske izmere praktično uporabni, mora biti del vsake izmere spremljanje njihovega pojavljanja, njihovo odkrivanje in odstranjevanje. V splošnem lahko rečemo, da med samo izmero lahko odkrijemo le grobe pogreške, ki izstopajo po velikosti.

Če imamo na razpolago nadštevilna opazovanja, je treba obravnavati opazovanja glede prisotnosti grobih pogreškov tudi pred izravnavo ter po izravnavi opazovanj v matematičnem modelu. Pred izravnavo opazovanj izvajamo analizo opazovanj na osnovi skladnosti opazovanj in pogojev, ki jih morajo ta izpolniti (izpolnjevanje pogojnih enačb). Predhodna ocena prisotnosti grobo pogrešenih opazovanj omogoča odkrivanje grobih pogreškov, ki so po velikosti manjši od pogreškov, ugotovljenih med samo izmero. Odkrivanje grobih pogreškov na osnovi rezultatov izravnave temelji na analizi popravkov opazovanj. Težava, ki nastane med izravnavo opazovanj, obremenjenih z grobimi pogreški, je v tem, da izravnava teži k porazdelitvi njihovega vpliva tudi na popravke drugih opazovanj.

Postopke, ki omogočajo odkrivanje grobo pogrešenih opazovanj, vključuje večina programske opreme, namenjene za izravnavo opazovanj v geodeziji. Članek naj bi vzpodbudil kakovostnejšo izrabo tovrstnih računalniških programov, saj se analizam rezultatov opravljenega dela zaradi nepoznavanja metod odkrivanja grobo pogrešenih opazovanj največkrat izognemo. V prispevku obravnavamo metode odkrivanja grobih pogreškov v opazovanjih na osnovi rezultatov izravnave v matematičnem modelu enostavne geodetske mreže.

1 MODEL GAUSS-MARKOVA (MGM)

MGM je linearen matematični model, ki ga sestavljajo funkcijske in stohastične povezave spremenljivk, vključenih v model. Povezuje slučajni vektor opazovanj **l** s slučajnim vektorjem neznank Δ.

Enačbe MGM-ja lahko zapišemo kot (Caspary, 1988; Kuang, 1996):

$$E(\mathbf{l}) = \mathbf{B}\boldsymbol{\Delta}\ , \qquad \text{(ali } \mathbf{v} + \mathbf{B}\boldsymbol{\Delta} = \mathbf{f}\text{)}$$

$$D(\mathbf{l}) = \mathbf{P}^{-1} \cdot \sigma_0^2 \tag{1.1}$$

ob morebitno prisotnih datumskih vezeh:

$$\mathbf{H}^{\mathrm{T}}\boldsymbol{\Delta} = \mathbf{0}\ . \tag{1.2}$$

Enačbi (1.1) podajata statistične lastnosti vektorja opazovanj **l**, zvezo med slučajnimi vektorji opazovanj **l**, popravkov opazovanj **v** in neznank Δ. Oznake in simboli imajo naslednji pomen:

$E(\cdot)$	pričakovana vrednost,	$D(\cdot)$	disperzija slučajnega vektorja,
$\mathbf{l}$	vektor opazovanj,	$\mathbf{f}$	vektor odstopanj,
$\mathbf{B}$	matrika koeficientov neznank,	$\mathbf{H}$	datumska matrika,
$\mathbf{v}$	vektor popravkov opazovanj,	Δ	vektor neznank,
σ_0^2	referenčna varianca a priori,	$\mathbf{P}$	matrika uteži vektorja opazovanj.

Predoločen sistem $\mathbf{v}+\mathbf{B\Delta}=\mathbf{f}$ lahko rešimo po metodi najmanjših kvadratov, ob kateri moramo izpolniti pogoj:

$$\mathbf{v}+\mathbf{B\Delta}=\mathbf{f} \quad \text{ob: } \Phi=\mathbf{v}^{\mathrm{T}}\mathbf{Pv}=\text{minimum} \quad \text{in} \quad \mathbf{H}^{\mathrm{T}}\mathbf{\Delta}=\mathbf{0}\,. \tag{1.3}$$

Rešitev je naslednja (Kuang, 1996):

$$\mathbf{N}=\mathbf{B}^{\mathrm{T}}\mathbf{PB}\,, \qquad \mathbf{\Delta}=(\mathbf{N}+\mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Pf}$$

$$\mathbf{R}=\mathbf{I}-\mathbf{B}(\mathbf{N}+\mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P}\,,\quad \mathbf{v}=\mathbf{Rf}\,, \tag{1.4}$$

$$\hat{\mathbf{l}}=\mathbf{l}+\mathbf{v}\,,$$

kjer je: $\mathbf{N}$ matrika koeficientov normalnih enačb,

$\mathbf{R}$ matrika nadštevilnosti,

$\hat{\mathbf{l}}$ vektor izravnanih opazovanj.

Referenčno varianco a posteriori izračunamo:

$$\hat{\sigma}_0^2=\frac{\mathbf{v}^{\mathrm{T}}\mathbf{Pv}}{r}\,, \tag{1.5}$$

kjer je r = n – n_0 število nadštevilnih opazovanj oziroma število prostostnih stopenj v matematičnem modelu; n–število opazovanj; n_0-minimalno število opazovanj, potrebnih za enolično rešitev problema. V matematičnem modelu nastopajo še naslednje količine:

$$\mathbf{Q}_{\Delta\Delta}=(\mathbf{N}+\mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{N}(\mathbf{N}+\mathbf{HH}^{\mathrm{T}})^{-1}$$ matrika kofaktorjev neznank,

$$\mathbf{Q}_{vv}=\mathbf{P}^{-1}-\mathbf{B}(\mathbf{N}+\mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}$$ matrika kofaktorjev popravkov opazovanj,

$$\mathbf{Q}_{\hat{l}\hat{l}}=\mathbf{B}(\mathbf{N}+\mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}$$ matrika kofaktorjev izravnanih opazovanj

in pripadajoče kovariančne matrike:

$$\mathbf{\Sigma}_{\Delta\Delta}=\sigma_0^2\mathbf{Q}_{\Delta\Delta}\,,\ \mathbf{\Sigma}_{vv}=\sigma_0^2\mathbf{Q}_{vv}\,,\ \mathbf{\Sigma}_{\hat{l}\hat{l}}=\sigma_0^2\mathbf{Q}_{\hat{l}\hat{l}}\,. \tag{1.6}$$

2 METODE ODKRIVANJA GROBIH POGREŠKOV

Osnova za vse postopke ugotavljanja prisotnosti grobih pogreškov pred izravnavo ali na osnovi rezultatov izravnave so statistični testi. Ti so povezani s porazdelitvijo verjetnosti popravkov opazovanj ter njihovo skladnostjo z znano ali predpostavljeno porazdelitvijo opazovanj.

Postopek statističnega testiranja hipotez lahko predstavimo z zaporedjem korakov (Kuang, 1996):

1. določitev ničelne hipoteze H_0 in alternativne hipoteze H_1;
2. določitev testne statistike T;
3. določitev porazdelitve verjetnosti za testno statistiko T pod H_0;
4. izbira stopnje značilnosti testa α;
5. izbira jakosti testa $1-\beta$;
6. izračun mejnih vrednosti kritičnega intervala;
7. odločitev glede zavrnitve H_0.

Pri testiranju hipotez se lahko pojavita dve napaki (Krakiwsky et al., 1999). **Napaka I. vrste** je zavrnitev H_0, ko je ta pravilna. Verjetnost, da storimo napako I. vrste, imenujemo stopnja značilnosti testa in jo označimo z α. Verjetnost, da je odločitev pravilna, ko ničelno hipotezo sprejmemo in je ta pravilna, imenujemo stopnja zaupanja in je enaka $1-\alpha$. **Napaka II. vrste** je opredeljena kot napaka, ki jo storimo, kadar sprejmemo H_0, ko je ta napačna. Verjetnost, da storimo napako II. vrste, označimo z β. Verjetnost, da zavrnemo H_0, ko je ta napačna, imenujemo jakost testa in jo označimo z $1-\beta$.

Kot ničelno hipotezo v splošnem postavimo naslednjo trditev:

H_0 : Model je pravilen in popoln; porazdelitvene predpostavke se ujemajo s stvarnostjo. (Med opazovanji ni grobih pogreškov.)

2.1 Globalni test modela

Po izravnavi izvedemo globalni test modela, s katerim testiramo skladnost referenčne variance a posteriori $\hat{\sigma}_0^2$ in referenčne variance a priori σ_0^2. Pod predpostavko ničelne hipoteze H_0 bi morali biti referenčni varianci statistično skladni.

Na tej podlagi tvorimo testno statistiko (Caspary, 1988):

$$T = \frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{\sigma_0^2} = \frac{r \cdot \hat{\sigma}_0^2}{\sigma_0^2}, \tag{2.1}$$

ki ima pod predpostavko ničelne hipoteze H_0 porazdelitev χ^2 z r prostostnimi stopnjami:

$$T \mid H_0 \sim \chi^2(r). \tag{2.2}$$

Pričakovana vrednost testne statistike je kar število prostostnih stopenj r oziroma število nadštevilnih opazovanj v modelu:

$$E(T \mid H_0) = r \ , \tag{2.3}$$

iz česar izhaja:

$$E\left(\hat{\sigma}_0^2 / \sigma_0^2 \mid H_0\right) = 1 \ \text{oziroma} \ E\left(\hat{\sigma}_0^2 / \sigma_0^2 \mid H_0\right) = \sigma_0^2 \tag{2.4}$$

Za stopnjo značilnosti globalnega testa modela običajno izberemo α = 0,05. Na podlagi izbrane vrednosti α izračunamo kritično vrednost $\chi^2_{1-\alpha/2}(r)$ testne statistike T. Testno statistiko nato primerjamo s kritično vrednostjo. Če velja:

$$T \leq \chi^2_{1-\alpha/2}(r) \ , \tag{2.5}$$

test ne kaže na statistično značilno protislovje med porazdelitvijo opazovanj in matematičnim modelom. V nasprotnem primeru, ko velja $T > \chi^2_{1-\alpha/2}(r)$, je treba raziskati, zakaj so ali model ali opazovanja ali uteži opazovanj nepravilni. Raziskavi lahko pripomorejo nadaljnji statistični testi.

Kadar H_0 zavrnemo, se v praksi omejimo na dve alternativni hipotezi (Kuang, 1996):

- H_{1-1}: nepravilne uteži opazovanj;
- H_{1-2}: prisotnost grobih pogreškov v opazovanjih.

Problem rešujemo postopoma. Najprej preverimo možnost, da smo opazovanjem dodelili napačne uteži. Če zavrnemo H_0 zaradi slabo ocenjene natančnosti opazovanj, so popravki še vedno normalno porazdeljeni. V primeru ko so popravki po svoji velikosti glede na natančnost uporabljenega instrumentarija primerno veliki in je referenčna varianca a posteriori $\hat{\sigma}_0^2$ prevelika glede na σ_0^2, je možen vzrok za zavrnitev H_0 slabo sestavljena kovariančna matrika opazovanj $\boldsymbol{\Sigma}$ in jo je treba

$$\tilde{\boldsymbol{\Sigma}} = \frac{\hat{\sigma}_0^2}{\sigma_0^2} \boldsymbol{\Sigma} \ . \tag{2.6}$$

Izravnavo ponovimo z uporabo kovariančne matrike $\tilde{\boldsymbol{\Sigma}}$. Ničelne hipoteze sedaj ne moremo zavrniti. Če pa so bili v prvem koraku izravnave prisotni tudi popravki opazovanj, ki so preveliki glede na natančnost uporabljenega instrumentarija, moramo opazovanja testirati glede na prisotnost grobih pogreškov v opazovanjih. Težava, ki se pri tem pojavi, je naslednja: s povečanjem vrednosti elementov kovariančne matrike zmanjšamo uteži opazovanj, kar oteži iskanje grobih pogreškov.

Ničelno hipotezo globalnega testa lahko zavrnemo tudi, kadar testna statistika ne doseže spodnje meje kritične vrednosti, ki jo izračunamo kot $\chi^2_{\alpha/2}(r)$. Če je:

$$T < \chi^2_{\alpha/2}(r)\,, \tag{2.7}$$

pomeni, da smo podcenili natančnost opazovanj, torej ima referenčna varianca a priori preveliko vrednost (opazovanja so natančnejša od naše predpostavke - σ_0^2).

2.2 Iskanje grobih pogreškov z metodo pregledovanja popravkov opazovanj Data Snooping

Metodo je leta 1968 predlagal Baarda. Osnovne predpostavke metode so naslednje (Ambrožič, 1996):

- opazovanja in popravki opazovanj so normalno porazdeljeni;
- zanesljivo poznamo referenčno varianco a priori σ_0^2 ;
- razen za eno grobo pogrešeno opazovanje velja model Gauss-Markova.

Data Snooping je sestavljena metoda. V prvem koraku z globalnim testom ugotavljamo eventualno prisotnost grobo pogrešenih opazovanj. V drugem koraku s pregledovanjem posameznih popravkov opazovanj lociramo in odstranimo grobo pogrešena opazovanja. Z Data Snoopingom lahko odkrijemo le eno grobo pogrešeno opazovanje naenkrat. Ostale grobe pogreške odkrivamo postopoma in s ponovitvijo postopka potem, ko smo izločili odkriti grobi pogrešek.

Ničelna hipoteza Data Snoopinga je naslednja:

H_0 : med opazovanji ni grobih pogreškov.

Določiti je treba tudi stopnjo značilnosti α_0 in stopnjo jakosti (1- β_0) enodimenzionalnega testa, s katerim bomo testirali posamezne popravke. Običajni vrednosti sta naslednji:

$\alpha_0 = 0{,}001$,

$\beta_0 = 0{,}20$.

Prisotnost grobega pogreška med opazovanji določimo z globalnim testom, s katerim hkrati preverimo vse popravke opazovanj. **Data Snooping predpostavlja, da zanesljivo poznamo referenčno varianco a priori**, zato v primeru zavrnitve ničelne hipoteze globalnega testa sprejmemo alternativno hipotezo H_1:

H_1: eno grobo pogrešeno opazovanje je povzročilo zavrnitev H_0.

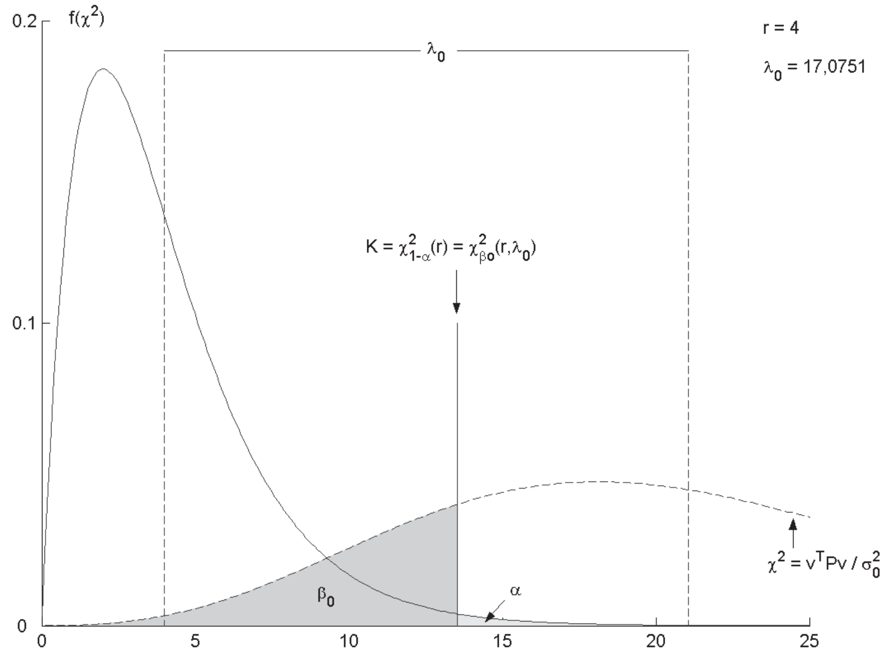
Če ničelno hipotezo zavrnemo, se testna statistika porazdeljuje po necentralni porazdelitvi χ^2 s parametrom necentralnosti λ in r prostostnimi stopnjami:

$$T \mid H_1 \sim \chi^2(r, \lambda) \tag{2.8}$$

Pri tem je parameter necentralnosti λ neposredno odvisen od vrednosti α_0 in β_0 ter r. λ predstavlja premik pričakovane vrednosti testne statistike, tako da testna statistika vzorca, ki vsebuje grobo pogrešeno opazovanje, preseže kritično vrednost z verjetnostjo 1-β_0. Ta verjetnost je opredeljena z enačbo (Caspary, 1988):

$$P = \left\{\chi^2(r,\lambda_0) > \chi^2_{1-\alpha}(r)\right\} = 1 - \beta_0 \tag{2.9}$$

in je prikazana na sliki 1 (vrednosti ustrezajo obravnavanemu primeru v razdelku 3).



Slika 1: *Funkcija gostote verjetnosti porazdelitve χ^2 pri danih α, β_0 in r:*

Polna linija: *centralna porazdelitev χ^2 za r = 4 in λ = 0.*

Črtkana linija: *necentralna porazdelitev χ^2 za r = 4 in λ_0 = 17,0751.*

Nekatere vrednosti $\sqrt{\lambda_0}$ za izbrana α_0, β_0 in r = 1 (predpostavljen je en grobi pogrešek) so podane v preglednici 1.

β_0	α_0							
	0,00001	0,00005	0,0001	0,0005	0,001	0,01	0,025	0,05
0,10	5,6987	5,3372	5,1721	4,7623	4,5721	3,8574	3,5230	3,2416
0,20	5,2588	4,8972	4,7322	4,3224	4,1322	3,4175	3,0830	2,8000
0,30	4,9416	4,5814	4,4161	4,0067	3,8165	3,1002	2,7658	2,4844

Preglednica 1: *Vrednosti $\sqrt{\lambda_0}$ za α_0, β_0 in r = 1.*

Pred izvedbo globalnega testa je treba izračunati njegovo stopnjo značilnosti α. Stopnja značilnosti α r-dimenzionalnega testa mora ustrezati enodimenzionalnem pregledu popravkov s stopnjo značilnosti α_0 in jakostjo $1-\beta_0$ tako, da grobi pogrešek povzroči zavrnitev ničelne hipoteze globalnega testa z verjetnostjo $1-\beta_0$. To dosežemo z uporabo istega parametra necentralnosti λ_0 za oba testa (Caspary, 1988):

$$\lambda_0 = \lambda(\alpha, \beta_0, r) = \lambda(\alpha_0, \beta_0, 1) \tag{2.10}$$

Zaradi implicitne oblike enačbe (2.10) se za pridobivanje vrednosti α raje uporabljajo nomogrami. Nekatere izračunane vrednosti α za β_0 = 0,20 prikazuje preglednica 2:

β_0= 0,20	α_0					
r	0,0001	0,0005	0,001	0,01	0,025	0,05
1	0,0001	0,0005	0,0010	0,0100	0,0250	0,0502
2	0,0003	0,0015	0,0028	0,0233	0,0524	0,0955
3	0,0007	0,0030	0,0055	0,0384	0,0796	0,1351
4	0,0013	0,0050	0,0089	0,0542	0,1054	0,1695
5	0,0021	0,0076	0,0130	0,0703	0,1296	0,1997
6	0,0031	0,0106	0,0177	0,0862	0,1520	0,2263
7	0,0044	0,0141	0,0229	0,1018	0,1728	0,2499
8	0,0058	0,0179	0,0284	0,1169	0,1921	0,2711
9	0,0075	0,0220	0,0343	0,1315	0,2100	0,2902
10	0,0094	0,0265	0,0404	0,1455	0,2266	0,3075
11	0,0115	0,0311	0,0467	0,1589	0,2421	0,3233
12	0,0138	0,0360	0,0532	0,1717	0,2565	0,3378
13	0,0162	0,0410	0,0598	0,1840	0,2700	0,3511
14	0,0188	0,0462	0,0664	0,1958	0,2827	0,3634
15	0,0216	0,0515	0,0731	0,2070	0,2946	0,3748

Preglednica 2: *Vrednosti α za α_0 in β_0 = 0,20 pri različnih prostostnih stopnjah* r.

Če zaradi prisotnosti grobega pogreška zavrnemo ničelno hipotezo, moramo za odkritje grobega pogreška pregledati posamezne popravke opazovanj. Za vsak popravek v_i izračunamo njegov standardizirani popravek (Caspary, 1988):

$$u_i = \frac{v_i}{\sigma_{v_i}} = \frac{v_i}{\sigma_0 \sqrt{q_{v_i v_i}}} \tag{2.11}$$

kjer je $q_{v_i v_i}$ diagonalni element matrike kofaktorjev popravkov opazovanj, ki pripada popravku v_i. Posamezne standardizirane popravke nato primerjamo s kritično vrednostjo pri izbrani stopnji značilnosti α_0. Standardizirani popravki u_i so standardizirano normalno porazdeljeni:

$$u_i \mid H_0 \sim N(0,1), \tag{2.12}$$

kar nam omogoča izračun kritične vrednosti u_{α_0}. Kritična vrednost standardizirane normalne porazdelitve je odvisna le od α_0.

Popravek označimo kot popravek verjetno grobo pogrešenega opazovanja v primeru, če (Caspary, 1988):

$$|u_i| > u_{1-\alpha_0/2} \tag{2.13}$$

Metoda omogoča tudi izračun najmanjšega grobega pogreška, ki ga bo test še prepoznal (Caspary, 1988):

$$\nabla_{0i} = \frac{\sigma_0}{p_i}\sqrt{\frac{\lambda_0}{q_{v_i v_i}}} \tag{2.14}$$

Če grobi pogrešek doseže vrednost ∇_{0i}, potem je verjetnost za odkritje tega pogreška enaka 1 - β_0, medtem ko je verjetnost, da zavržemo dobro opazovanje, enaka α_0.

Izračunamo lahko tudi faktor k_{0i}, ki nam pove, kolikokrat mora grobi pogrešek preseči standardno deviacijo opazovanja, da ga bo mogoče odkriti z verjetnostjo 1-β_0 (Caspary, 1988):

$$k_{0i} = \sqrt{\frac{\lambda_0}{p_i \cdot q_{v_i v_i}}} \; . \tag{2.15}$$

Iz enačb (2.14) in (2.15) lahko neposredno izračunamo velikost grobega pogreška, ki ga lahko odkrijemo z metodo Data Snooping. Ob predpostavki, da so uteži opazovanj zanesljivo poznane, je velikost grobega pogreška, ki ga še lahko odkrijemo, odvisna le od parametra necentralnosti λ_0. Če izračunamo:

$$\sqrt{\lambda_0} = \frac{\nabla_{0i} \cdot p_i \cdot \sqrt{q_{v_i v_i}}}{\sigma_0} \qquad \text{ali} \qquad \sqrt{\lambda_0} = k_{0i} \cdot \sqrt{p_i \cdot q_{v_i v_i}} \; , \tag{2.16}$$

lahko iz preglednice 1 odčitamo stopnjo značilnosti in stopnjo jakosti testa za odkrivanje želene velikosti grobega pogreška.

2.3 Lociranje grobih pogreškov

Ko smo odkrili popravke verjetno grobo pogrešenih opazovanj, je naslednji korak lociranje grobih pogreškov. To pomeni, da je treba ugotoviti, ali so odkriti popravki verjetno grobo pogrešenih opazovanj v_i posledica grobih pogreškov v opazovanjih l_i. V splošnem je uspešnost lociranja grobih pogreškov odvisna od geometrije matematičnega modela in od števila in velikosti grobih pogreškov v opazovanjih. Splošna zahteva je, da mora biti natančnost opazovanj med seboj podobna – primerljiva ter da mora biti število grobih pogreškov v opazovanjih manjše, kot je skupna nadštevilnost sistema.

Ker imamo opravka z grobimi pogreški majhne velikosti, je pravilno lociranje verjetno grobo pogrešenih opazovanj močno odvisno od geometrijskih lastnosti problema. V primeru da med opazovanji obstaja samo eno grobo pogrešeno opazovanje, lahko zaradi koreliranosti popravkov opazovanj popravek verjetno grobo pogrešenega opazovanja v_i kaže na prisotnost grobega pogreška v opazovanju samo, če ima to opazovanje dominantno število nadštevilnosti r_{ii}. Število nadštevilnosti posameznih opazovanj predstavljajo diagonalni členi matrike nadštevilnosti **R** (Kuang, 1996; Seemkooei, 2001):

$$\mathbf{R} = \mathbf{I} - \mathbf{B}(\mathbf{N} + \mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} \ . \qquad (1.5)$$

Število nadštevilnosti r_{ii} je dominantno, če velja:

$$r_{ii} > \max\left\{ |r_{ji}|;\ j = 1..n;\ j \neq i \right\}, \qquad (2.17)$$

pri čemer so r_{ji} izvendiagonalni členi matrike nadštevilnosti **R**. V obratnem primeru ($r_{ii} \leq \max\{|r_{ji}|; j = 1..n; j \neq i\}$) je ugotavljanje grobo pogrešenih opazovanj oteženo. Z drugimi besedami, grobi pogrešek v kateremkoli opazovanju l_j ($j = 1..n; j \neq i$) lahko povzroči prevelik popravek v_i opazovanja l_i.

Število nadštevilnosti je v celoti odvisno od geometrije problema in neodvisno od dejanskih opazovanj. To pomeni, da ga lahko določimo vnaprej in tako poskrbimo za enakomerno kakovost modela na celotnem območju.

2.4 Test τ

Data Snooping lahko izvajamo samo, kadar zanesljivo poznamo referenčno varianco a priori σ_0^2. Te pa v praksi v večini primerov ne poznamo dovolj zanesljivo. V tem primeru grobe pogreške odkrivamo s testom τ, ki ga je leta 1976 predlagal Pope (Popeova metoda). Ostale predpostavke so pri testu τ enake kot pri metodi Data Snooping.

Testna statistika T_i testa τ se pod predpostavko ničelne hipoteze porazdeljuje po porazdelitvi τ z r prostostnimi stopnjami (Caspary, 1988):

$$T_i = \frac{|v_i|}{\hat{\sigma}_{v_i}} = \frac{|v_i|}{\hat{\sigma}_0 \sqrt{q_{v_i v_i}}} \sim \tau(r) \ . \qquad (2.18)$$

Za stopnjo značilnosti testa, ki je sestavljen iz n posameznih testov, običajno izberemo $\alpha = 5$ %. Natančen izračun stopnje značilnosti α_0 za n enodimenzionalnih testov žal ni mogoč, saj so popravki, in zato posledično tudi testi, statistično odvisni.

Približna enačba za izračun α_0 je naslednja (Krakiwsky et al., 1999):

$$\alpha_0 \approx 1 - (1 - \alpha)^{1/n} \ . \qquad (2.19)$$

Preglednice porazdelitve τ so težje dostopne od preglednic porazdelitve t. Pri pretvorbi kritičnih vrednosti nam pomaga naslednja enačba (Kuang, 1996):

$$\tau_{1-\alpha_0/2}(r) = \frac{\sqrt{r}\, t_{1-\alpha_0/2}(r-1)}{\sqrt{r-1+t^2_{1-\alpha_0/2}(r-1)}} \,. \tag{2.20}$$

Test uporabimo zaporedoma za vse standardizirane popravke T_i. Ničelno hipotezo zavrnemo, če posamezna vrednost testne statistike T_i popravka v_i presega kritično vrednost $\tau_{1-\alpha_0/2}$. Če velja:

$$T_i > \tau_{1-\alpha_0/2} \,, \tag{2.21}$$

je nakazana možnost, da pripada popravek v_i verjetno grobo pogrešenemu opazovanju.

Test τ tako kot Data Snooping temelji na predpostavki, da je le eno opazovanje v MGM-ju obremenjeno z grobim pogreškom. Če opazovanja vključujejo več grobih pogreškov, je priporočen naslednji pragmatični pristop, za katerega pa ni nujno, da bo dal ustrezne rezultate. Opazovanje z največjo testno statistiko izpustimo in ponovimo izravnavo z n−1 opazovanji, kar nam da nove popravke in novo oceno za a posteriori referenčno varianco $\hat{\sigma}_0^2$. Izračunamo novo vrednost za α_0 in ponovimo test z r−1 prostostnimi stopnjami. Postopek ponavljamo, dokler niso označeni vsi verjetni grobi pogreški.

Ko so nekateri popravki označeni kot preveliki, je postopek lociranja in odstranitve grobih pogreškov iz opazovanj enak kot pri postopku Data Snooping.

Težava pri uporabi testa τ je v tem, da je referenčna varianca a posteriori $\hat{\sigma}_0^2$ obremenjena s prisotnostjo grobih pogreškov v opazovanjih. Večji kot so grobi pogreški, večja bo vrednost $\hat{\sigma}_0^2$, ki bo na ta način zmanjševala vrednost testne statistike T_i. Ob uporabi testa τ tako lahko nekateri manjši grobi pogreški ostanejo neodkriti.

2.5 Danska metoda

Osnovna zamisel danske metode je v tem, da večji popravki pripadajo manj natančnim opazovanjem, in obratno. Pri oceni parametrov v MGM s pomočjo izravnave po metodi najmanjših kvadratov zato zamenjamo a priori uteži z novimi, ki so funkcija popravkov opazovanj. Nova izravnava nam poda nove popravke, iz katerih ponovno izračunamo nove uteži. Postopek ponavljamo do konvergence (dokler ni razlika med zaporednimi utežmi manjša od želene vrednosti). Postopek običajno zahteva 5 do 10 iteracij. Danska metoda ne sloni na statistični teoriji. Ne potrebuje nobenih predpostavk glede stohastičnih lastnosti opazovanj, zato tudi ni treba izvajati nobenih statističnih testov.

Za izračun novih vrednosti uteži je bilo predlaganih več funkcij. (Leick, 1995) predlaga naslednjo funkcijo za izračun novih uteži:

$$p_{i+1} = p_i f(v_i), i = 1..n; \qquad f(v_i) = \begin{cases} 1; & \text{za } |v_i| < c \cdot \sigma_i, \\ \exp\left(-\dfrac{|v_i|}{c \cdot \sigma_i}\right); & \text{sicer.} \end{cases} \qquad (2.22)$$

kjer je σ_i standardna deviacija opazovanja l_i. Vrednost konstante c običajno izberemo med 2 in 3, odvisno od nadštevilnosti in kakovosti podatkov opazovanj. Cilj metode je zmanjšanje vpliva grobih pogreškov na oceno neznank in zagotovitev rezultatov, skladnih s pričakovanji.

Po končani izravnavi primerjamo končne vrednosti uteži z a priori utežmi. Bistveno spremenjene uteži posameznih opazovanj nakazujejo možna grobo pogrešena opazovanja (po izvedbi danske metode je utež, ki pripada grobo pogrešenemu opazovanju, največkrat blizu 0). Ustrezna opazovanja je zato treba natančno pregledati in po možnosti popraviti. Očitna grobo pogrešena opazovanja odstranimo in jih po potrebi ponovimo.

Za končno oceno neznanih parametrov obstajata dve možnosti:

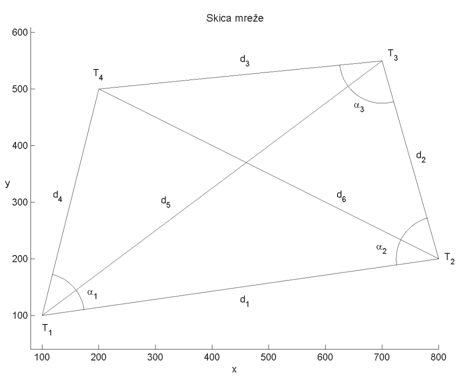
(i) po odstranitvi vseh grobih pogreškov ponovimo izravnavo z a priori utežmi;

(ii) kot najboljšo oceno parametrov privzamemo parametre iz zadnje iteracije in obdržimo vsa opazovanja.

V nadaljevanju ilustriramo opisane metode na praktičnem primeru.

3 RAČUNSKI PRIMER ISKANJA GROBIH POGREŠKOV

Ugotavljanje prisotnosti grobih pogreškov v nadaljevanju predstavljamo na primeru geodetskega četverokotnika, v katerem je bilo izmerjenih 6 dolžin in trije horizontalni koti (slika 2). Opazovanja so izravnana s posredno izravnavo v prosti mreži. Datum mreže določajo notranje vezi med koordinatami točk.



Slika 2: *Skica geodetske mreže.*

Približne koordinate točk [m]: $T_1^0 = (100{,}100)$ $T_3^0 = (700{,}550)$

$T_2^0 = (800{,}200)$ $T_4^0 = (200{,}500)$

Merjene dolžine: $d_1 = 707{,}1415$ m Merjeni koti: $\alpha_1 = 67^0\ 50'\ 07{,}7''$

$d_2 = 364{,}0075$ m $\alpha_2 = 82^0\ 10'\ 47{,}9''$

$d_3 = 502{,}5091$ m $\alpha_3 = 100^0\ 14'\ 18{,}6''$

$d_4 = 412{,}3003$ m

$d_5 = 750{,}0058$ m

$d_6 = 670{,}8538$ m

Standardne deviacije: $\sigma_d = 5\text{mm} + 5\text{ppm}$ $\sigma_\alpha = 10''$

d_3 je bila »pokvarjena« za 8-kratno vrednost svoje standardne deviacije: $d_3 = 502{,}5692$ m.

3.1 Opredelitev geodetskega datuma z notranjimi vezmi

Geodetski datum zagotavlja položaj, orientacijo in merilo geodetske mreže. V primeru 2D-mreže morajo notranje vezi zagotoviti, da se ne premakne težišče mreže (položaj: dva datumska parametra), da se mreža ne zasuka (orientacija: en parameter) in da ostane povprečna razdalja med težiščem in posameznimi točkami nespremenjena (merilo: en parameter). Ker imamo v našem primeru merjene tudi dolžine, te že določajo merilo mreže. Potrebujemo le tri datumske parametre. Prvi dve zahtevi (položaj in orientacija mreže) zapišemo v matrični obliki kot (Kuang, 1996):

$$\mathbf{H}^{\mathrm{T}}\boldsymbol{\Delta} = \mathbf{0} \tag{3.1}$$

Datumska matrika $\mathbf{H}^{\mathrm{T}}$ ima obliko:

$$\mathbf{H}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 1 \\ -y_1^0 & x_1^0 & -y_2^0 & x_2^0 & \cdots & -y_m^0 & x_m^0 \end{bmatrix}, \tag{3.2}$$

kjer so x_i^0, y_i^0 približne koordinate točk mreže (v matematičnem ravninskem koordinatnem sistemu).

3.2 Posredna izravnava

Enačbe opazovanj:

Enačba, ki povezuje opazovano dolžino med točkama T_i in T_j z neznankami v izravnavi:

$$F_{dij}: \quad d_{ij} - \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} = 0. \tag{3.3}$$

Enačba, ki povezuje opazovani horizontalni kot na točki T_i proti točkam T_j in T_k z neznankami v izravnavi:

$$F\alpha_{jik}: \quad \alpha_{jik} - \operatorname{atan}\frac{y_j - y_i}{x_j - x_i} + \operatorname{atan}\frac{y_k - y_i}{x_k - x_i} = 0. \tag{3.4}$$

Linearizirane enačbe popravkov dolžinskih opazovanj:

$$v_{dij} + \frac{\Delta x_{ij}^0}{d_{ij}^0}\delta x_i + \frac{\Delta y_{ij}^0}{d_{ij}^0}\delta y_i - \frac{\Delta x_{ij}^0}{d_{ij}^0}\delta x_j - \frac{\Delta y_{ij}^0}{d_{ij}^0}\delta y_j = d_{ij}^0 - d_{ij} \tag{3.5}$$

Linearizirane enačbe popravkov kotnih opazovanj:

$$v_{\alpha jik} + \left(\frac{\Delta y_{ik}^0}{(d_{ik}^0)^2} - \frac{\Delta y_{ij}^0}{(d_{ij}^0)^2}\right)\delta x_i + \left(\frac{\Delta x_{ij}^0}{(d_{ij}^0)^2} - \frac{\Delta x_{ik}^0}{(d_{ik}^0)^2}\right)\delta y_i +$$

$$+\frac{\Delta y_{ij}^0}{(d_{ij}^0)^2}\delta x_j - \frac{\Delta x_{ij}^0}{(d_{ij}^0)^2}\delta y_j - \frac{\Delta y_{ik}^0}{(d_{ik}^0)^2}\delta x_k + \frac{\Delta x_{ik}^0}{(d_{ik}^0)^2}\delta y_k = \alpha_{jik}^0 - \alpha_{jik} \tag{3.6}$$

Opazovanja izravnamo s posredno izravnavo proste mreže, kot je opisano v razdelku 1 (1.4–1.7).

3.2.1 Data Snooping – rezultati odkrivanja grobih pogreškov

Referenčna varianca a priori σ_0^2:	σ_0^2 = 0,00001	
Izbrana stopnja značilnosti testa α_0:	α_0 = 0,001	
Izbrana jakost testa 1 - β_0:	β_0 = 0,20	
Parameter necentralnosti λ_0:	λ_0 = 17,0751	(iz preglednice 1)
Število prostostnih stopenj r:	r = 4	
Stopnja značilnosti r-dimenzionalnega testa α:	α = 0,0089	(iz preglednice 2)

Globalni test modela

Testna statistika T (2.1):	T = 17,0185
Kritična vrednost $\chi^2_{1-\alpha}(r)$:	$\chi^2_{1-0,0089}(4)$ = 13,5445

Ničelna hipoteza globalnega testa modela je zavrnjena ($T > \chi^2_{1-\alpha}(r)$).

Pregled popravkov opazovanj

Kritična vrednost standardizirane normalne porazdelitve $u_{\alpha_0=0,001}$: $u_{1-\alpha_0/2} = 3,2905$

Vrednosti standardiziranih popravkov u_i prikazuje preglednica 3.

Opazovanje	$\lvert u_i \rvert$ (2.11)	∇_{0i} (2.14)	k_{0i} (2.15)
d_1	1,0080	0,0686 m	8,0379
d_2	3,3115	0,0909 m	13,3315
d_3	4,1142*	0,0574 m	7,6447
d_4	2,8442	0,0993 m	14,0680
d_5	2,1549	0,0536 m	6,1251
d_6	3,3765	0,0549 m	6,5657
α_1	0,9483	0,000218	4,5018
α_2	1,4601	0,000221	4,5550
α_3	1,1066	0,000232	4,7893

Preglednica 3: *Standardizirani popravki, najmanjši prepoznavni grobi pogreški in faktor* k *pri Data Snoopingu.*

Standardizirani popravek opazovanja d_3 najbolj presega kritično vrednost!
Število nadštevilnosti meritve d_3 (element matrike **R**): $r_{3,3} = 0,2922$
Izvendiagonalni elementi matrike **R** (1.4):

$r_{ji\,(j\neq i)}$ = [0,0736; 0,1249; 0,1007; -0,2331; -0,2957; -0,0009; -0,0010; 0,0008]

Verjeten grobi pogrešek! Ker $r_{3,3} \approx |r_{6,3}|$, moramo biti pri izločanju opazovanja previdni!

3.2.2 Test τ – rezultati odkrivanja grobih pogreškov

Izbrana stopnja značilnosti n-dimenzionalnega testa α: $\alpha = 0,05$

Stopnja značilnosti enodimenzionalnega testa α_0 (2.19): $\alpha_0 = 0,0057$

Kritična vrednost testa (2.20): $\tau_{1-\alpha_0/2}(r) = 1,9435$

Testne statistike opazovanj T_i (2.18):

Opazovanje	T_i		Opazovanje	T_i
d_1	0,4887		α_1	0,4598
d_2	1,6054		α_2	0,7079
d_3	1,9946*		α_3	0,5365
d_4	1,3789			
d_5	1,0447			
d_6	1,6369			

Preglednica 4: *Vrednosti testnih statistik pri testu τ.*

Testna statistika opazovanja d_3 je presegla kritično vrednost. Možen grobi pogrešek! Matrika **R** je enaka kot v primeru Data Snoopinga.

3.2.3 Danska metoda – rezultati odkrivanja grobih pogreškov

Konstanta c: 2

Konvergenčni pogoj: $p_{i+1} - p_i < 10^{-6}$; i = 1..n

Število potrebnih iteracij: 6

Pregled uteži:

Opazovanje	a priori uteži (P_1)	a posteriori uteži (P_6)
d_1	0,137253	0,137253
d_2	0,214994	0,214994
d_3	0,177185	0,00000001*
d_4	0,200542	0,200542
d_5	0,130611	0,130611
d_6	0,143279	0,0141214
α_1	4254,517030	4254,517030
α_2	4254,517030	4254,517030
α_3	4254,517030	4254,517030

Preglednica 5: *Uteži v prvi in šesti iteraciji danske metode.*

Utež opazovanja d_3 je bistveno spremenjena. Verjeten grobi pogrešek!

4 ZAKLJUČEK

Odkrivanje grobih pogreškov znotraj same izravnave izkorišča nadštevilnost in celotno geometrijo matematičnega modela. Zato je še posebej občutljivo na manjše grobe pogreške. Iskanje grobih pogreškov se zato izvaja le v primerih, ko je odkrita njihova morebitna prisotnost na osnovi popravkov opazovanj, ki nimajo pričakovane porazdelitve verjetnosti. Takšni popravki opazovanj (angl. outliers) so najboljši namig, kje iskati problematična opazovanja. S tem se lahko izognemo nepotrebnemu in neorganiziranemu iskanju po celotnem nizu opazovanj.

Pomembno je razumeti, da pri pojavu popravka, ki ni zadovoljil posameznega testa, še ne pomeni, da gre v opazovanju za grobi pogrešek. Opazovanje le označimo, ga preverimo in sprejmemo odločitev o tem, ali ga bomo obdržali ali zavrgli. Izpuščanje opazovanj »na slepo« ni nikoli priporočljivo. Grobi pogrešek enega opazovanja ima običajno učinek tudi na popravke drugih opazovanj, zato se lahko samo na osnovi rezultatov statističnega testa poleg grobo pogrešenega opazovanja kot sumljiva označijo tudi druga opazovanja. Če je med opazovanji označeno kot verjetno grobo pogrešeno eno ali več opazovanj, se prične iskanje morebiti prisotnega grobega pogreška. Tudi v primerih, ko test označi samo eno opazovanje kot grobo pogrešeno, se lahko zgodi, da nam izravnava z izpuščenim označenim opazovanjem ne bo dala zadovoljivih rezultatov. V tem primeru je treba opazovanje vrniti v vektor opazovanj in iskati grobi pogrešek drugje.

Literatura in viri:

Ambrožič, T. (1996). Ocena stabilnosti točk v geodetski mreži. Magistrska naloga. Ljubljana: Fakulteta za gradbeništvo in geodezijo, Oddelek za geodezijo.

Caspary, W. F. (1988). Concepts of Network and Deformation Analysis. School of Surveying, The University of New South Wales, Kensington, N.S.W., Australia.

Krakiwsky, E. J., Craymer, M. R., Szabo, D. J., Vanicek, P. (1999). Development and Testing of In-Context Confidence Regions for Geodetic Survey Networks. Spletna stran, datum dostopa: 6. 10. 2003, (URL): www2.geod.nrcan.gc.ca/~craymer/pubs/incontext99.pdf

Kuang, S. (1996). Geodetic Network Analysis and Optimal Design: Concepts and Applications, Ann Arbor Press, Inc. ZDA.

Leick, A. (1995). GPS Satellite Surveying. John Willey & Sons, Canada.

Seemkooei, A. A. (2001). Comparison of Reliability and Geometrical Strength Criteria in Geodetic Networks. Journal of Geodesy, št. 75, str. 227–233.

mag. Dejan Grigillo, univ. dipl. inž. geod.
FGG – Oddelek za geodezijo, Jamova 2, SI-1000 Ljubljana
e-pošta: dgrigill@fgg.uni-lj.si

izr. prof. dr. Bojan Stopar, univ. dipl. inž. geod.
FGG – Oddelek za geodezijo, Jamova 2, SI-1000 Ljubljana
e-pošta: bstopar@fgg.uni-lj.si

Prispelo v objavo: 6. november 2003