

Efektivitas Penggunaan Sampel berdasarkan Pelenyapan Data: Kasus Pemboran pada Kawasan Emas Vein

Nur Ali Amri

Teknik Pertambangan, FTM, Universitas Pembangunan Nasional "Veteran" Yogyakarta
Jl. SWK 104 (Lingkar Utara) Condongcatur Yogyakarta Indonesia 55283
nuraliamri@upnyk.co.id

Abstrak. Tulisan ini membahas efektivitas penggunaan data berdasarkan estimasi Jackknife-kriging menggunakan simulasi pelenyapan sampel data pada satu kawasan vein tambang emas. Estimasi berdasarkan sepuluh, dua puluh, tiga puluh dan empat puluh persen pelenyapan data awal menggunakan *leave one out*. Kualitas perhitungan dinilai berdasarkan beberapa ralat dan homogenitas distribusi merujuk kepada koefisien variasi. Berdasarkan perhitungan didapat bahwa, estimasi mengacu sepuluh dan dua puluh persen pelenyapan data masih cukup efektif digunakan.

Kata kunci: Sampel data, emas vein, kriging Jackknife, pelenyapan data, ralat.

Abstract. This paper discusses the effectiveness of data based on Jackknife-kriging estimation using simulation of gold data sampling in vein region. The estimations based on ten, twenty, thirty and forty percent of leave-one-out of initial losing data. The quality judged by some errors and also the homogeneity of the distribution refers to the coefficient of variation. Based on the calculations, it is found that the estimation referring to ten and twenty percent of data elimination are still quite effective.

Keywords: Data sample, gold-vein, Jackknife kriging, losing data, error.

A. Latar belakang

Penggunaan jumlah sampel, terutama dalam statistik spasial berpengaruh pada kualitas hasil perhitungan. Sampel yang terlalu sedikit menyebabkan hasil perhitungan statistik yang kurang – bahkan – tidak baik. Sedangkan jumlah sampel terlalu banyak (terlebih sampel yang diperoleh dengan biaya tinggi semisal, pemboran pada kawasan yang relatif sulit dijangkau), meskipun menghasilkan perhitungan lebih representatif, tetapi secara praktik mengakibatkan pemborosan baik secara finansial, tenaga maupun waktu [1]. Berdasarkan pertimbangan ini, upaya perhitungan menggunakan kaedah geostatistik, terutama estimasi kriging berdasar simulasi pelenyapan data dikerjakan.

B. Material dan Metode

Lokasi kajian berada pada kawasan Ciurug, Pongkor, milik PT. UBPE Pongkor, Kecamatan Nanggung, Bogor, Indonesia dengan koordinat $6^{\circ}39'$ Lintang Selatan dan $106^{\circ}33'$ Bujur Timur.

Teknik kriging lazim digunakan pada estimasi data spasial, tidak terkecuali dalam pertambangan

logam [2, 3]. Pelaksanaan validasi silang pelenyapan berdasarkan formula Jackknife [4, 5]. Katakanlah sampel s_k ($i=1, 2, \dots, k, \dots, n$) merupakan k data yang akan dilenyapkan, maka validasi silang kriging berdasarkan *leave one out* adalah

$$Z(s_k) = \sum_{s_i \in G \setminus \{s_k\}} w_i Z(s_i) \quad (1)$$

$$\text{dimana, } \sum_{i=1}^n w_i = 1.$$

Sedangkan semi-variogram *robust* berdasarkan formula [6], yaitu

$$\gamma(h) = \frac{\left[\frac{1}{|N(h)|} \sum_{i,j \in N(h)} (Z(s_i) - Z(s_j))^2 \right]^{1/4}}{0,914 + [0,988/|N(h)|]} \quad (2)$$

Akurasi perhitungan validasi silang pada estimasi pelenyapan data menggunakan indikator median deviasi mutlak [7, 8]. Nilai koefisien variasi bagi masing-masing estimasi berdasar pelenyapan data digunakan untuk mengetahui heterogenitas data [9]. Ketelitian menggunakan

median deviasi mutlak (MDM), rata-rata ralat mutlak (RRM), sedangkan homogenitas data menggunakan koefisien variasi (KV) sebagaimana [1, 10]. Beberapa formula yang digunakan adalah:

- Median deviasi mutlak gabungan sampel random ke- i adalah

$$\varepsilon_{md_i} = \text{median}_{\mathbf{s}_{ik} \in G_s} |\widehat{Z}(\mathbf{s}_{ik}) - Z(\mathbf{s}_{ik})|, \quad (3)$$

dimana median deviasi mutlaknya adalah,

$$\varepsilon_{md} = \text{median}_j \{|\widehat{Z}(\mathbf{s}_j) - Z(\mathbf{s}_j)|\}.$$

- Rata-rata ralat mutlak adalah

$$\bar{\varepsilon}_{rm} = \frac{1}{n} \sum_{i=1}^n |\widehat{Z}(\mathbf{s}_i) - Z(\mathbf{s}_i)|. \quad (4)$$

- Koefisien variasi adalah

$$KV = \frac{s^2}{\bar{Z}(\mathbf{s})}, \quad (5)$$

s^2 adalah varians dengan rata-rata estimasi adalah

$$\bar{Z}(\mathbf{s}) = \frac{1}{n} \sum_{i=1}^n Z(\mathbf{s}_i). \quad (6)$$

Dua fitting digunakan dalam kajian ini ialah, *robust* semi-variogram dengan *weighted least squared* basis eksponensial dan sferis [11, 5, 12, 13]. Estimator menggunakan ordinary kriging. Metode validasi silang menggunakan pengasingan M titik. Sebagai pelaksanaan, M titik dipilih secara random dari himpunan S . Selanjutnya, nilai bijih emas pada titik diestimasi dibandingkan dengan nilai sebenarnya. Proses ini diulang untuk beberapa kali dengan gabungan sampel M persen yang berbeda.

C. Hasil dan Analisis

Parameter semivariogram mendasarkan kepada dua model fitting ialah, berdasarkan semivariogram empiris *robust* basis *weighted least squares* sferis (RWLSS) dan eksponensial (RWLSE). Selengkapnya parameter tersebut disajikan pada Tabel 1. Sebagaimana kolom dua, basis eksponensial tidak didapat nilai nugget. Meski demikian seperti tampak pada kolom empat, kedua model menghasilkan sill yang

hampir sama, masing-masing ialah 11,102 dan 11,391. Panjang range (praktikal) basis eksponensial lebih panjang dibandingkan basis sferis.

Kalkulasi ordinary kriging mengacu 6150 (atau titik dengan jarak teratur 150×41) didapat nilai-nilai minimum, rata-rata, maksimum dan median deviasi mutlak estimasi sebagaimana Tabel 2. Dengan interval 11,935, model RWLSS menghasilkan rata-rata estimasi sebesar 5,098. Nilai estimasi ini sedikit lebih besar dibandingkan dengan estimasi berdasar model RWLSE (iaitu, 4,958). Sedangkan berdasar RWLSE didapat nilai median deviasi mutlak lebih kecil dibanding basis model RWLSS.

Tabel 1. Parameter semivariogram data asal berdasar dua model

Model	Nugget	C_1	Sill	Range
RWLSS	1,40	9,69	11,10	149,54
RWLSE	0,00	11,39	11,39	179,88

Tabel 2. Statistik nilai kriging berdasar grid 150×41 titik data asal

Model	Minimum	Rata-rata	Maksimum	MDM
RWLSS	1,45	5,10	13,39	1,64
RWLSE	1,23	4,958	14,28	1,55

Pengujian berdasarkan *jack-knifing* dengan pelenyapan 10, 20, 30 dan 40 persen data asal, di mana setiap pelenyapan dilakukan 200 kali pengulangan. Pelenyapan sepuluh persen terhadap data asal maknanya, perhitungan mendasarkan kepada 115 data. Pelenyapan dua puluh persen dari data asal berarti, terdapat 102 data yang digunakan dalam perhitungan. Demikian seterusnya. Sebagaimana Tabel 3, Tabel 4 dan Gambar 1, tiga nilai ialah, median deviasi mutlak (MDM), rata-rata ralat mutlak (RRM) dan Koefisien Variasi (KV) disajikan. Terlihat dari ketiganya berlaku kenaikan sejalan dengan pelenyapan data. Sebagaimana Tabel 2, model fitting berdasar semivariogram empiris *robust* basis *weighted least squares* eksponensial menghasilkan MDM yang secara umum lebih kecil berbanding model berdasar semivariogram empiris *robust* basis *weighted least squares* sferis. Kasus serupa juga berlaku untuk RRM dan KV,

sebagaimana ditampilkan masing-masing pada Tabel 3 dan Tabel 4.

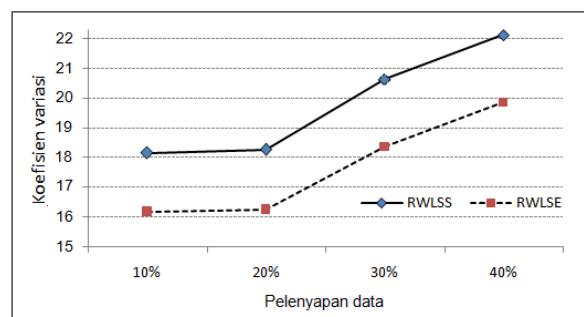
Tabel 3. Median deviasi mutlak berdasar pelenyapan 10, 20, 30 dan 40 persen data asal

Model	10%	20%	30%	40%
RWLSS	1,790	1,822	1,877	1,911
RWLSE	1,715	1,789	1,832	1,871

Tabel 4. Rata-rata ralat mutlak berdasar pelenyapan 10, 20, 30 dan 40 persen data asal

Model	10%	20%	30%	40%
RWLSS	2,130	2,199	2,227	2,259
RWLSE	2,080	2,152	2,172	2,205

Di dalam kaedah statistik, koefisien variasi dapat digunakan sebagai pendekripsi homogenitas/heterogenitas distribusi data. Sebagaimana Tabel 4, pada 10 dan 20 persen pelenyapan tidak terjadi perubahan nilai yang signifikan. Perubahan dimulai pada 30 persen pelenyapan. Mengacu perubahan nilai sebagaimana tabel koefisien variasi, KV diperoleh informasi bahwa, data masih efisien digunakan hingga pengurangan 13 data, bahkan dimungkinkan sampai 102.



Gambar 1. Koefisien variasi berdasar pelenyapan 10, 20, 30 dan 40 persen data asal

Pengecekan menggunakan 115 sampel data dimulai dengan perhitungan variografi. Nilai-nilai parameter semovariogram berdasar 115 data dipaparkan pada Tabel 5. Baris satu adalah paparan parameter berdasar *fitting* RWLSS, sedangkan baris dua adalah sajian parameter basis eksponensial. Sebagaimana yang berlaku pada *fitting* data asal, nilai-nilai *sill* dan *range* untuk pelenyapan data juga berlaku demikian, di mana basis model RWLSE menghasilkan *sill* dan *range* yang lebih besar dibandingkan model RWLSS.

Tabel 5. Parameter semivariogram 115 data berdasar dua model

Model	Nugget	C ₁	Sill	Range
RWLSS	0	10,292	10,29	151,00
RWLSE	0	10,617	10,62	195,12

Estimasi menggunakan parameter kedua-dua basis dipaparkan dalam Tabel 6. Nilai-minimum perhitungan kriging ditampilkan pada kolom dua. Kolom tiga adalah paparan rata-rata estimasi kriging berdasar 6150 titik acuan. Sedangkan kolom empat dan kolom lima, masing-masing adalah tampilan nilai-nilai maksimum dan *median absolute deviation* estimasi. Sebagaimana kolom tiga baris pertama, rata-rata yang dihasilkan basis RWLSS ialah, 5,146. Nilai ini sedikit lebih besar berbanding rata-rata basis RWLSE. Jika dikomparasi dengan perhitungan kriging terhadap data asal, rata-rata mengacu pelenyapan data sedikit lebih besar dengan pautan 0,048. Sedangkan perhitungan estimasi berdasar pelenyapan data basis RWLSE menghasilkan rata-rata 5,006, sedikit lebih besar dibanding rata-rata kriging data asal, iaitu. 4,958; atau juga terpaut 0,048.

Tabel 6. Statistik nilai kriging berdasar grid 150×41 titik 115 data

Model	Minimum	Rata-rata	Maksimum	MDM
RWLSS	1,45	5,15	13,05	1,48
RWLSE	1,24	5,00	14,05	1,50

D. Kesimpulan

Berdasarkan uraian di atas dapat disimpulkan bahwa

- Tulisan ini menyajikan estimasi ordinary kriging, yang pada perjalannya dilakukan menggunakan simulasi pelenyapan data.
- Akurasi didasarkan pada median deviasi mutlak dan rata-rata ralat mutlak.
- Efektivitas penggunaan sampel data pada distribusi emas vein, terutama yang diperoleh melalui pemboran dapat diketahui berdasarkan simulasi pelenyapan data.
- Dari 128 data tersedia, perhitungan estimasi sebenarnya masih bisa dilakukan dengan pengurangan menjadi 115, bahkan 102 sampel data.

E. Ucapan Terimakasih

Terimakasih penulis sampaikan kepada manajemen PT. Aneka Tambang (Tbk), Unit Bisnis Penambangan Emas (UBPE) Pongkor, yang telah memberikan berbagai-bagai data terkait penelitian ini. Terimakasih juga disampaikan kepada Universitas Pembangunan Nasional "Veteran" Yogyakarta yang telah memberi kesempatan kepada penulis untuk melakukan penelitian.

F. Daftar Pustaka

- [1] Dunn, O.J. & Clark, V.A. 2009. *Basic Statistics; A primer for the biomedical sciences*. John Wiley & Sons, Inc. Fourth edition New Jersey.
- [2] Champigny, N. & Armstrong, M. 1993. Geostatistics for the estimation of gold deposits, *International Journal of Mineral Deposit* 28: 279-282.
- [3] Roy, D., Butt, S.D. & Frempong, P.K. 2004. Geostatistical resource estimation for the Poura narrow-vein gold deposit. *CIM Bulletin ProQuest Science Journals*, Feb. (1077): 47-51.
- [4] Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence* 2: 1137–1143.
- [5] Quenouille, M.H. 1949. Approximate tests of correlation in time series. *Journal R Stat. Soc. B* 11: 18 – 84.
- [6] Cressie, N. & Hawkins, D.M. 1980. Robust estimation of the variogram. *International Journal of Association for Mathematical Geology*, 12 (2): 115-125.
- [7] Amri, N.A., Jemain, A.A & Hassan, W.F.W. 2014. Ordinary kriging based on OLS-WLS fitting semivariogram: case of gold vein precipitation. The 3rd International Conference on Mathematical Sciences. 1039-1045.
- [8] Liu, X.H. & Liu, Y. 2008. The accuracy assessment in areal interpolation: An empirical investigation. *Journal of Science: Technological Sciences* 51 (1): 62-71.
- [9] Amri, N.A., Abdul Aziz, J. & Fudholi, A. 2016. Optimization of spatial data sample for gold mineral prediction. *Journal of Asian Research Publishing Network* 11(15): 9065–9068.
- [10] James, J., Witten, D., Hastie, T. & Tibshirani, R. 2013. *An introduction to statistical learning with an application in R*. Springer Science+Business Media New York.
- [11] Bocci, C. & Dabrowski, A.R. 2002. Fitting data to linear combination of covariograms. *Journal of Environmental Modeling and Assessment* 7: 39-46.
- [12] Cressie, N. 1993. *Statistics for spatial data*. New York. Chichester. Toronto. Brisbane. Singapore: John Wiley & Sons, Inc.
- [13] Schaffrin, B. & Uzun, S. 2008. Surface reconstruction via total least-squares adjustment of the semi-variogram. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. (37): 141-148.