

A Study on Accuracy of Paddy Harvest Area Estimation on Area Sampling Frame Method*

Mulianto Raharjo¹, Anang Kurnia^{2‡}, and Hari Wijayanto³

^{1,2,3}Department of Statistics, IPB University, Indonesia
[‡]corresponding author: anangk@apps.ipb.ac.id

Copyright © 2021 Mulianto Raharjo, Anang Kurnia, and Hari Wijayanto. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

There was unsynchronized national rice data until 2017, which indicating that influenced by the differences in calculation methods between government agencies. The Indonesian Central Bureau of Statistics (BPS Statistics), the most responsible agency for national rice data, collected rice plant areas data using the paddy statistical assessment method (SP-Padi). Subjective elements from various parties potentially influenced the result of this assessment method. The development of a new method to overcome this matter has been started by the government since 1993. In 2018 the method, which is named the Area Sample Frame (ASF) method, was officially used by the government under the coordination of BPS. The ASF method divides the area into grids: blocks, segments, and sub-segments. This new method has several issues related to the methodology used in determining the sampling method. This study was conducted to evaluate the accuracy of paddy harvest area estimation on the ASF method through a sampling simulation process of the ASF method with various scenarios. With 20 simulated scenario combinations, it was found that the difference percentage average between the harvested area of the population and the harvested area of the sample to the sub-district area was 0.00062%, and the mean square error (MSE) was 0.0041%. So it can be concluded that the ASF methodology is an unbiased method and is good enough to accommodate various strata diversity in any region.

Keywords: area sampling frame, BPS Statistics, paddy harvest area, simulation.

* Received: Jun 2021; Reviewed: Nov 2021; Published: May 2022

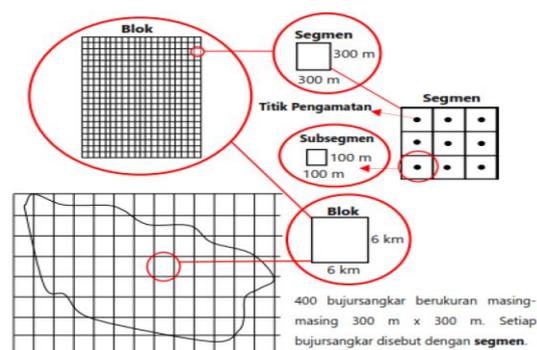
1. Pendahuluan

Kementerian Pertanian menyebutkan produksi beras nasional tahun 2018 sebesar 46,5 juta ton, sedangkan konsumsi beras nasional sebesar 33,47 juta ton sehingga terdapat surplus beras sebesar 13,03 juta ton hingga akhir tahun 2018. Di lain pihak, Badan Pusat Statistik (BPS) menyebutkan produksi beras nasional sebesar 32,42 juta ton, sedangkan konsumsi beras nasional sebesar 29,57 juta ton sehingga surplus beras diperkirakan sebesar 2,85 juta ton. Sementara itu, Kementerian Perdagangan memutuskan untuk melakukan impor beras sebanyak dua juta ton selama tahun 2018, namun kebijakan impor ini ditolak Perusahaan Umum Badan Urusan Logistik (Perum Bulog). Hal tersebut menunjukkan terjadinya perbedaan pada data beras nasional tahun 2018.

Ketidaksinkronan kebijakan yang dikeluarkan antar instansi tersebut diduga disebabkan adanya perbedaan metode perhitungan yang dilakukan masing-masing instansi sehingga muncul perbedaan angka jumlah produksi beras nasional. BPS selaku instansi yang paling bertanggung jawab terhadap data beras nasional, melakukan pengumpulan data luas tanaman padi menggunakan metode penaksiran statistik pertanian padi (SP-Padi) yang dilakukan pada tingkat kecamatan. Metode penaksiran yang dimaksud diperoleh dengan beberapa cara, yaitu menggunakan sistem blok pengairan, laporan petani kepada kepala desa, banyaknya benih yang digunakan, dan *eye estimate* (pandangan mata) berdasarkan luas baku (Badan Pusat Statistik, 2018). Dengan metode seperti itu, akan terdapat unsur subjektif dari enumerator di lapangan dalam mencatat angka luas tanaman padi sehingga metode pengumpulan data ini tidak memenuhi salah satu kaidah statistika.

Kerangka sampel area (KSA) merupakan metode pengumpulan data luas panen yang dikembangkan oleh BPS bersama Badan Pengkajian dan Penerapan Teknologi (BPPT) untuk memperbaiki metode yang sebelumnya agar lebih objektif dengan melibatkan teknologi (Badan Pusat Statistik, 2018). Metode ini juga menjadi salah satu solusi untuk permasalahan subjektivitas dari metode penaksiran SP-Padi. Pemerintah melalui Kantor Staf Presiden (KSP) pada tahun 2018 menetapkan pengukuran luas panen padi nasional menggunakan metode KSA.

Metode KSA membagi wilayah ke dalam blok, segmen, dan subsegmen untuk kemudian dilakukan penarikan contoh acak. Blok berukuran 6 km x 6 km dan segmen berukuran 300 m x 300 m dibentuk pada kerangka sampel sawah, sehingga terdapat 400 segmen per blok. Setiap segmen memiliki stratifikasi lahan sawah. Kemudian dilakukan pembagian pada setiap segmen menjadi subsegmen yang berukuran 100 m x 100 m. Ilustrasi pembagian area menjadi blok, segmen, dan subsegmen terdapat pada Gambar 1.



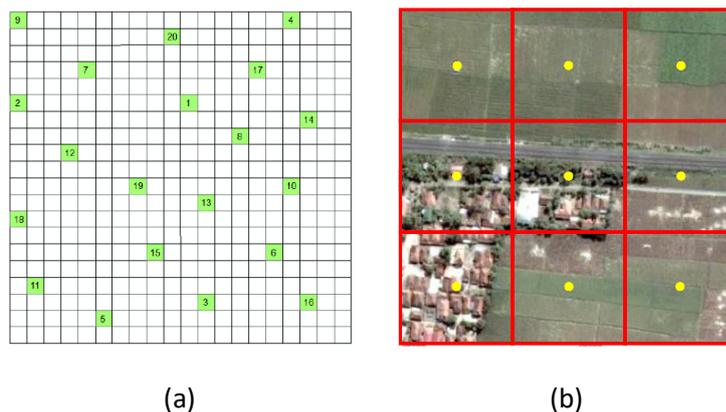
Gambar 1: Ilustrasi pembagian area menjadi blok, segmen, dan subsegmen

Stratifikasi yang dibentuk terbagi menjadi empat kelompok besar, yaitu poligon-poligon bukan persawahan (Strata-0/S0), poligon-poligon persawahan irigasi (Strata-1/S1), poligon-poligon sawah non-irigasi/tadah hujan (Strata-2/S2), dan poligon-poligon lahan kering untuk tanaman pangan yang biasa disebut dengan tegalan (Strata-3/S3). Strata-0 tidak akan dialokasikan sebagai sampel segmen karena dianggap tidak ada unsur penggunaan lahan untuk persawahan sekaligus untuk mengurangi jumlah sampel. Satuan pengukuran yang digunakan pada setiap subsegmen KSA adalah kategori fase pertumbuhan tanaman padi dan penggunaan lahan yang lain. Kategori fase pertumbuhan padi tersebut terbagi menjadi delapan kategori, yaitu vegetatif awal, vegetatif akhir, generatif, panen, persiapan lahan, puso, sawah bukan padi, dan bukan sawah (Badan Pusat Statistik, 2018).

Segmen pada setiap blok dipilih secara acak sebanyak 5% sehingga terdapat 20 sampel segmen. Ilustrasi pemilihan sampel segmen terdapat pada Gambar 2(a). Lalu pada setiap sampel segmen terpilih dilakukan pengamatan kategori fase pertumbuhan padi pada sembilan titik pusat subsegmen. Jika sulit menjangkau titik pusat subsegmen, enumerator dapat melakukan pengamatan di sekitar titik pusat dengan batas radius 10 m dari titik pusat subsegmen. Ilustrasi titik pengamatan segmen pada titik pusat subsegmen terdapat pada Gambar 2(b).

Pengamatan dilakukan secara berulang setiap bulannya (tujuh hari terakhir di bulan pengamatan) sepanjang tahun menggunakan telepon genggam bersistem operasi android melalui foto dan input data manual. Foto yang diunggah digunakan untuk validasi fase pertumbuhan padi yang diinput enumerator. Hasil input data fase pertumbuhan padi digunakan sebagai luas panen padi pada bulan tersebut dan prediksi luas panen padi bulan-bulan selanjutnya.

Pengembangan metode baru ini memiliki beberapa isu terkait metodologi yang digunakan dalam penentuan metode penarikan contoh (*sampling*). Isu pertama, yaitu penyusunan blok, segmen, dan subsegmen yang berbentuk *grid* (segi empat). Kedua,



Gambar 2: Ilustrasi (a) pemilihan 5% sampel segmen pada blok dan (b) titik pengamatan segmen pada titik pusat subsegmen

pengambilan 5% segmen 300 m x 300 m pada blok 6 km x 6 km terpilih. Ketiga, pengambilan sembilan titik (satu titik pada setiap subsegmen) pada segmen terpilih.

Ketiga isu tersebut tentu saja akan memengaruhi luaran yang dihasilkan dari KSA, yaitu luas panen. Ketidaktepatan metode penarikan contoh yang digunakan akan mengakibatkan rendahnya akurasi nilai luas panen yang dihasilkan. Perlu dilakukan pemeriksaan akurasi dari metode penarikan contoh pada KSA. Maka dari itu, tujuan

dari penelitian ini adalah melakukan kajian evaluasi akurasi luas panen padi pada metode KSA melalui proses simulasi penarikan contoh metode KSA dengan berbagai skenario.

2. Metodologi

2.1 Data

Data yang digunakan merupakan data simulasi kategori fase tumbuh padi dan strata per segmen. Skenario simulasi ditentukan secara subjektif. Kategori fase tumbuh padi yang dibangkitkan adalah pada petak berukuran 20 m x 20 m, sehingga pada setiap subsegmen akan terdapat 25 petak. Ilustrasi pembagian petakan pada setiap subsegmen terdapat pada Gambar 3. Setiap petak diasumsikan memiliki kondisi fase tumbuh yang homogen sehingga setiap petak hanya memiliki satu nilai fase tumbuh. Sementara strata per segmen yang dibangkitkan adalah sesuai yang ditetapkan pada KSA, yaitu S1, S2, dan S3. Penelitian ini memfokuskan ke nilai luas panen pada setiap bulan pengamatan, sehingga fase tumbuh yang akan dibangkitkan bersifat biner, yaitu panen (1) dan tidak panen (0). Skenario pembangkitan data dijabarkan lebih detil pada Tabel 1 dan Tabel 2. Sebaran strata segmen per blok untuk setiap skenario ditentukan berdasarkan keragaman jumlah sampel segmen KSA tahun 2018 per provinsi (Badan Pusat Statistik, 2018) yang dikelompokkan menjadi 4 kelompok. Skenario keragaman proporsi strata pada segmen terpilih dan skenario keragaman proporsi status panen petakan pada subsegmen akan dikombinasikan, sehingga akan ada 20 kombinasi skenario yang akan dibangkitkan.

Tabel 2: Skenario keragaman proporsi strata pada segmen terpilih yang dibangkitkan

Skenario	Strata segmen terpilih per blok			
	17 S1, 2 S2, 1 S3 ^a	11 S1, 7 S2, 2 S3 ^b	7 S1, 11 S2, 2 S3 ^c	2 S1, 13 S2, 5 S3 ^d
A1	40%	30%	20%	10%
A2	10%	40%	30%	20%
A3	20%	10%	40%	30%
A4	30%	20%	10%	40%

^a17 segmen S1, 2 segmen S2, 1 segmen S3

^b11 segmen S1, 7 segmen S2, 2 segmen S3

^c7 segmen S1, 11 segmen S2, 2 segmen S3

^d2 segmen S1, 13 segmen S2, 5 segmen S3

Tabel 1: Skenario keragaman proporsi status panen petakan pada subsegmen yang dibangkitkan

Skenario	Status panen petakan per subsegmen					
	25 panen, 0 tidak ^a	20 panen, 5 tidak ^b	15 panen, 10 tidak ^c	10 panen, 15 tidak ^d	5 panen, 20 tidak ^e	0 panen, 25 tidak ^f
B1	20%	15%	15%	15%	15%	20%
B2	30%	15%	5%	5%	15%	30%
B3	5%	15%	30%	30%	15%	5%
B4	30%	30%	15%	15%	5%	5%
B5	5%	5%	15%	15%	30%	30%

^a25 petak panen, 0 petak tidak panen

^b20 petak panen, 5 petak tidak panen

^c15 petak panen, 10 petak tidak panen

^d10 petak panen, 15 petak tidak panen

^e5 petak panen, 20 petak tidak panen

^f0 petak panen, 25 petak tidak panen

Tabel 4: Daftar peubah data Potensi Desa tahun 2005 yang digunakan

Nomor peubah	Nama peubah	Keterangan
R1001	Luas desa/kelurahan	Total luas
R1002a	Luas lahan sawah berpengairan yang diusahakan	S1
R1002b	Luas lahan sawah tidak berpengairan yang diusahakan	S2
R1002c	Luas lahan sawah sementara tidak diusahakan	S0
R1003a	Lahan pertanian (kolam/tambak/perkebunan/hutan rakyat/padang rumput)	S0
R1003b	Ladang yang diusahakan	S3
R1003c	Ladang yang tidak diusahakan	S0
R1003d	Lahan untuk non pertanian (pemukiman/perumahan/pertokoan/industri dan lainnya)	S0

Untuk mendukung proses simulasi yang dilakukan, digunakan juga data Potensi Desa (Podes) Provinsi Jawa Barat tahun 2005. Data Podes ini digunakan untuk menghitung luas lahan yang panen dan tidak panen per kategori stratifikasi. Daftar peubah Podes 2005 yang digunakan terdapat pada Tabel 3, yaitu peubah-peubah yang terdapat pada subbab "Penggunaan Lahan". Meskipun data tersebut berselang belasan tahun dengan penelitian ini, namun tetap digunakan dengan pertimbangan bahwa pada data Podes tahun 2005 peubah-peubah pada subbab "Penggunaan Lahan" masih didiseminasikan. Selain itu, pada penelitian ini tidak menitikberatkan pada agregasi di tingkat kota/kabupaten, melainkan pada keragaman kondisi lahan yang ada. Dengan alasan yang sama, kondisi bahwa adanya pemekaran kabupaten di Jawa Barat, yaitu Kabupaten Bandung Barat dan Kabupaten Pangandaran, juga tidak menjadi suatu kendala.

Tabel 3: Tabel konversi kecamatan menjadi blok berdasarkan luas wilayah

Rentang konversi luas wilayah	Blok konversi	Banyaknya kecamatan
18–54 km ²	1 blok	225
54–90 km ²	2 blok	227
90–126 km ²	3 blok	83
126–162 km ²	4 blok	22

2.2 Metode Penelitian

Tahapan kajian yang dilakukan adalah sebagai berikut:

1. Melakukan praproses data untuk merapikan data luas wilayah dan luas sawah Podes tahun 2005 yang digunakan. Praproses data yang digunakan adalah *winsorizing*.
2. Melakukan konversi wilayah kecamatan menjadi blok. Konversi ini mengikuti ketentuan pada Tabel 4.
3. Melakukan penggerombolan blok berdasarkan peubah luas area per strata hasil konversi pada langkah (2).
4. Melakukan pembangkitan skenario keragaman keadaan lahan sawah sebanyak 1000 blok dengan mengikuti ketentuan 20 kombinasi skenario dari Tabel 1 dan Tabel 2.
5. Melakukan penarikan contoh acak (*random sampling*) blok berdasarkan hasil penggerombolan pada langkah (3).

6. Melakukan perhitungan agregasi luas lahan di tingkat kecamatan sesuai dengan metode KSA berikut (Badan Pusat Statistik, 2018):

- a. Menghitung rata-rata proporsi luas panen per strata

$$\bar{p}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} p_{hi}$$

Keterangan:

p_{hi} : proporsi luas panen terhadap total luas segmen i pada strata h

\bar{p}_h : rata-rata proporsi luas panen terhadap total luas segmen pada strata h

n_h : banyaknya sampel segmen pada strata h

- b. Menghitung dugaan total luas panen padi di suatu kecamatan

$$A = \sum_{h=1}^H D_h \bar{p}_h$$

Keterangan:

A : luas panen padi

D_h : luas wilayah pada strata h

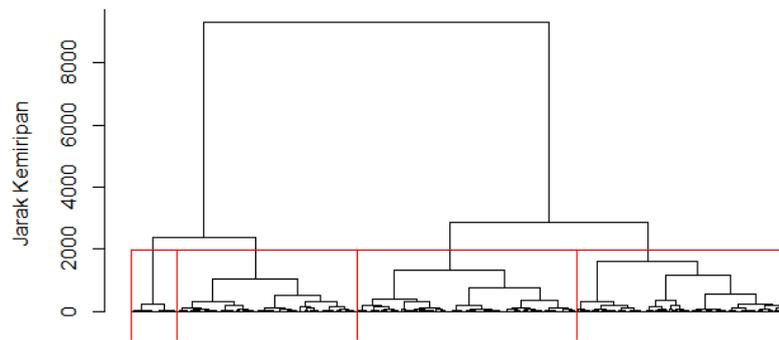
7. Menghitung selisih antara luas panen yang dibangkitkan dan hasil pendugaan luas panen dari contoh yang terambil.
8. Melakukan pengulangan langkah (4) sampai (7) sebanyak 300 kali.
9. Menghitung beberapa metrik untuk mengevaluasi besaran selisih pendugaan luas panen dari 300 kali ulangan, yaitu rata-rata, simpangan baku, statistik lima serangkai, *mean square error* (MSE), *mean absolute percentage error* (MAPE), dan *coefficient of variation* (CV).
10. Membuat diagram kotak garis dari selisih dugaan luas panen.

3. Hasil dan Pembahasan

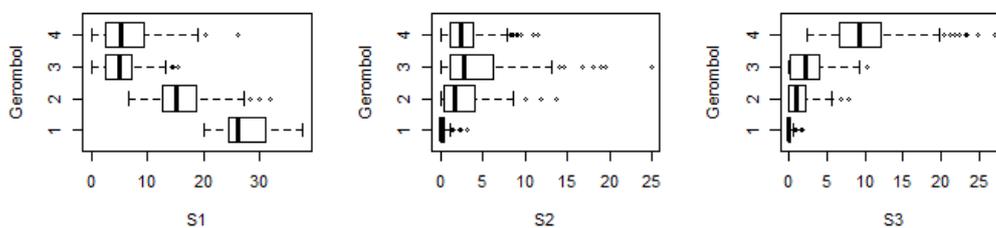
Simulasi dilakukan dengan mengacu pada keragaman luas sawah per kecamatan yang ada pada data Potensi Desa (Podes) tahun 2005, sedangkan *raw data* pada Podes tahun 2005 adalah desa/kelurahan. Selain itu, pada data tersebut terdapat beberapa nilai luas sawah yang agak ekstrem terlalu kecil ataupun terlalu luas. Sehingga perlu dilakukan praproses data *winsorizing* (Huber, 2011). Proses ini dilakukan dengan menganggap bahwa luas wilayah desa/kelurahan yang wajar adalah 5 sampai 80 km². Selanjutnya luas lahan per kategori yang ada pada Tabel 3 disesuaikan secara proporsional. Total terdapat 5808 desa/kelurahan dan 592 kecamatan di data Podes tahun 2005, 3755 desa diantaranya memiliki luas wilayah kurang dari 5 km² dan enam desa diantaranya memiliki luas wilayah lebih dari 80 km².

Blok KSA berukuran 6 km x 6 km sehingga memiliki luas 36 km² (Badan Pusat Statistik, 2018). Karena tidak dalam penelitian ini tidak menggunakan area spasial antar wilayah kecamatan, maka dalam tahapan simulasi digunakan pengelompokan luas wilayah kecamatan dibagi 36 km². Konversi luas wilayah menjadi banyaknya blok per kecamatan tersedia pada Tabel 4. Sementara 35 kecamatan yang luasnya kurang dari 18 km² dan lebih dari 162 km² tidak dimasukkan dalam proses simulasi. Proses konversi ini juga dilakukan untuk memudahkan perhitungan agregasi, yaitu menganggap setiap kecamatan hanya terdiri atas satu blok. Dengan demikian, setelah dikonversi menjadi blok terdapat

1016 blok yang digunakan dalam proses simulasi.



Gambar 4: Dendrogram hasil penggerombolan setiap blok hasil konversi



Gambar 3 Diagram kotak garis perbandingan luas wilayah per strata pada setiap gerombol

Tabel 5 Rataan luas wilayah per strata pada setiap gerombol

Gerombol	Luas S1 (km ²)	Luas S2 (km ²)	Luas S3 (km ²)	Banyaknya blok
1	27,50	0,37	0,20	71
2	15,80	2,43	1,40	278
3	4,89	4,10	2,49	338
4	6,17	2,87	10,10	329

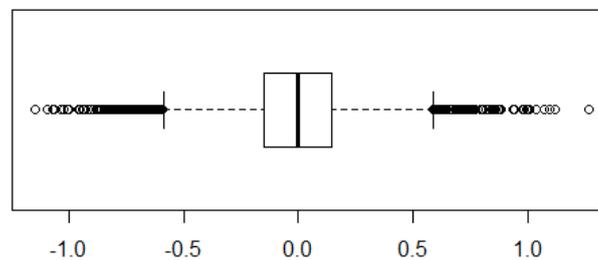
Untuk menentukan strata pada setiap blok, dilakukan proses penggerombolan. Penggerombolan merupakan salah satu teknik dalam *unsupervised classification* yang mengelompokkan objek berdasarkan kedekatan karakteristiknya (Johnson R. A, 2002; Sumertajaya dan Mattjik, 2011). Pautan pada penggerombolan yang dipilih adalah pautan Ward (Ward Jr, 1963) karena dalam pautan ini seringkali menghasilkan penggerombolan yang lebih baik dibanding pautan lain. Penggerombolan dilakukan dengan menggunakan peubah luas area per strata hasil konversi berdasarkan Tabel 3. Gambar 4 menunjukkan dendrogram hasil penggerombolan, terdapat empat gerombol yang dihasilkan. Statistik deskriptif dari keempat gerombol dapat terlihat pada Gambar 5 dan Tabel 5.

Blok yang termasuk dalam gerombol 1 mayoritas memiliki area Strata-1 yang paling luas dibandingkan gerombol lainnya, sedangkan area Strata-2 dan Strata-3 sangat sempit dibandingkan gerombol lainnya. Kemudian blok yang termasuk dalam gerombol 2 memiliki karakteristik yang hampir mirip dengan gerombol, namun area Strata-1 tidak seluas gerombol 1 serta area Strata-2 dan Strata-3 tidak sesempit gerombol 1. Lalu pada gerombol 3 yang terlihat mencolok berbeda

dibandingkan gerombol lainnya adalah luas area Strata-2 yang cenderung agak besar. Selanjutnya, pada gerombol 4 luas area Strata-3 lebih besar dibandingkan gerombol lainnya.

Dengan demikian, hasil penggerombolan dapat dicocokkan dengan kondisi sebaran strata per blok yang disimulasikan sesuai Tabel 1. Blok di gerombol 1 digunakan untuk melakukan simulasi kondisi pertama, yaitu 17 S1, 2 S2, dan 1 S3. Blok di gerombol 2 digunakan untuk melakukan simulasi kondisi kedua, yaitu 11 S1, 7 S2, dan 2 S3. Kemudian blok di gerombol 3 digunakan untuk melakukan simulasi kondisi ketiga, yaitu 7 S1, 11 S2, dan 2 S3. Terakhir, blok di gerombol 4 digunakan untuk melakukan simulasi kondisi keempat, yaitu 11 S1, 7 S2, dan 2 S3. Setelah didapatkan populasi blok untuk disimulasikan, selanjutnya dilakukan proses simulasi percontohan acak serta perhitungan agregasi prediksi luas panen.

Gambar 6 menunjukkan diagram kotak garis selisih antara luas panen populasi dan luas panen contoh dari setiap ulangan dibagi luas lahan per kecamatan.



Gambar 5: Diagram kotak garis selisih antara luas panen populasi dan luas panen contoh dari setiap ulangan dibagi luas lahan per kecamatan

Tabel 6 Statistik deskriptif agregasi selisih antara luas panen populasi dan luas panen contoh dari seluruh ulangan

Metrik	Nilai	Metrik	Nilai
Rataan	0,0076 km ²	Minimum	-74,4978 km ²
Simpangan baku	13,1	Kuartil 1	-5,9739 km ²
MSE	171	Median	0,1332 km ²
MAPE	0,374%	Kuartil 3	5,8946 km ²
CV	1.724	Maksimum	72,6599 km ²

Pada gambar tersebut terlihat bahwa rentang selisihnya tidak melebihi 75 km² dengan median 0,1332 km². Sedangkan hasil agregasi dari 300 ulangan terdapat pada Tabel 6. Rataan selisih antara luas panen populasi dan luas panen contoh dari seluruh ulangan adalah sebesar 0,0076 km² atau 0,76 ha. Nilai *mean absolute percentage error* (MAPE) yang dihasilkan adalah 0,374%. Dengan rata-rata selisih dan MAPE tersebut, maka dapat diambil kesimpulan bahwa metodologi KSA yang dilakukan oleh BPS sudah relevan dengan berbagai keragaman kondisi sebaran strata pada setiap daerah.

4. Simpulan

Evaluasi terhadap metodologi KSA dilakukan melalui proses simulasi pada dua hal, yaitu keragaman proporsi strata pada segmen terpilih dan keragaman proporsi status panen petakan pada subsegmen, dengan total terdapat 20 kombinasi skenario yang

dibangkitkan. Berdasarkan proses simulasi berbagai kondisi keheterogenan strata yang ada di lapangan, didapatkan rata-rata selisih antara luas panen populasi dan luas panen contoh sebesar 0,0076 km² dan nilai MAPE sebesar 0,374%. Dengan rata-rata selisih dan MAPE tersebut, maka dapat diambil kesimpulan bahwa metodologi KSA yang dilakukan oleh BPS merupakan metode yang tak bias dan sudah cukup baik mengakomodasi berbagai keragaman kondisi sebaran strata pada setiap daerah.

References

- [BPS] Badan Pusat Statistik. (2018). *Luas Panen dan Produksi Beras di Indonesia 2018*. Jakarta(ID): BPS.
- Huber, P. J. (2011). Robust statistics. In *International encyclopedia of statistical science* (pp. 1248–1251). Springer.
- Johnson R. A, W., D. W. (2002). *Applied multivariate statistical analysis* (8th ed., Vol. 5). Upper Saddle River, NJ: Prentice hall.
- Sumertajaya, I., & Mattjik, A. A. (2011). *Sidik Peubah Ganda Dengan Menggunakan SAS*. IPB PRESS.
- Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301): 236–244.