ISBN : 978.602.361.002.0

DESAIN SAMPLING UNTUK PEMODELAN SPATIAL

Bertho Tantular
Departemen Statistika FMIPA Universitas Padjadjaran
berthotantular@gmail.com

ABSTRAK. Desain sampling bergantung pada banyak faktor diantaranya: kondisi populasi, alat analisis yang digunakan, ketersediaan unit, waktu dan biaya. Dalam pemodelan statistika desain sampling bergantung pada model yang digunakan dalam analisis. Kondisi populasi yang heterogen dan memiliki dependensi spatial mengakibatkan asumsi saling bebas dan berdistribusi identik antar unit tidak terpenuhi, sehingga model yang digunakan adalah model spatial. Dengan demikian desain sampling harus mempertimbangkan adanya autokorelasi spatial. Selain itu penentuan ukuran sampel dalam desain sampling spatial harus mempertimbangkan adanya dependensi spatial antar unit analisis. Dalam pemodelan *conditional auto regressive* (CAR) maupun *spatial auto regressive* (SAR), matriks bobot spatial diperhitungkan dalam penentuan ukuran sampel menggunakan ukuran sampel efektif (*Effective Sample Size*). Dalam penentuan ukuran sampel efektif tersebut dibutuhkan informasi awal mengenai struktur korelasi spatialnya. Informasi tersebut diperoleh dari penelitian sebelumnya atau melalui *pilot study*. Kondisi populasi yang heterogen dan adanya dependensi mengakibatkan banyak konfigurasi yang berbeda pada setiap wilayah cacah sehingga sampling stratifikasi merupakan metode yang paling cocok untuk data spatial.

Kata Kunci: *Metode Sampling; Ukuran Sampel Efektif; Pemodelan Spatial*

1. PENDAHULUAN

Dalam penelitian survei desain sampling merupakan hal yang sangat menentukan hasil. Desain sampling dalam survei meliputi kegiatan sebagai berikut: menetapkan tujuan, menentukan sumber dan kendala, menentukan populasi sasaran, menyusun kerangka sampling, menentukan ukuran sampel dan menentukan metode sampling. Salah satu hal penting dalam suatu desain sampling adalahmenentukan ukuran sampel. Dalam menentukan ukuran sampel banyak faktor yang harus dipertimbangkan. Diantaranya adalah apakah penelitian tersebut bertujuan menaksir atau menguji hipotesis, teknik sampling apa yang digunakan, berapa banyak parameter yang terlibat, seberapa besar kekeliruan yang dapat ditoleransi dan berapa biaya yang harus dikeluarkan untuk penelitian tersebut. Selain itu metode sampling yang akan digunakan dalam survei memegang peranan sangat penting dalam penelitian mulai dari menentukan ukuran sampel, analisis hingga penarikan kesimpulan(Lohr [5]). Setiap teknik sampling mempunyai kelebihan dan kekurangannya masing-masing, akan tetapi semakin sederhana teknik sampling yang digunakan maka akan semakin sederhana rumus-rumus yang diperoleh. Meskipun demikian kita tidak dapat memaksakan teknik sampling yang sederhana untuk permasalahan yang rumit. Sehingga tetap diperlukan rumus-rumus untuk teknik sampling tersebut.

ISBN : 978.602.361.002.0

Untuk kondisi populasi yang heterogen dan memiliki dependensi spatial mengakibatkan asumsi saling bebas dan berdistribusi identik antar unit sampling tidak terpenuhi, sehingga desain sampling harus mempertimbangkan adanya dependensi spatial tersebut. Demikian pula dalam penentuan ukuran sampel harus mempertimbangkan adanya dependensi spatial antar unit analisis. Menurut Wang dkk. [8] analisis yang digunakan dapat didasarkan pada desain sampling atau didasarkan pada model. Umumnya analisis yang didasarkan pada desain sampling dengan menambahkan bobot sampling dalam perhitungannya. Sedangkan untuk analisis yang didasarkan pada model haus ditunkan dari asumsi yang mendasari model tersebut seperti pada model spatial. Vallejos dan Osario[7] mengusulkan metode penentuan ukuran sampel dalam model-model spatial seperti model*conditional auto regressive* (CAR) danmodel *spatial auto regressive* (SAR). Dalam penelitian ini akan diterapkan metode yang diusulkan oleh Wang dkk [8] dan Vallejos dan Osario[7] dalam penelitian mengenai Pendeteksian Faktor Resiko serta Pemetaan Penyebaran Tuberkolosis Anak Kecamatan Ngamprah Kabupaten Bandung dengan Pendekatan Model Multilevel dengan Efek Spasial (Pontoh dkk [6]).

2. METODE PENELITIAN

Tujuan dari metode sampling adalah untuk mendapatkan hasil yang lebih berkualitas dengan biaya lebih rendah (Wang dkk., [8]). Untuk itu diperlukan desain sampling yang baik. Dalam desain sampling ada dua hal yang harus ditentukan secara tepat yaitu ukuran sampel dan teknik sampling yang digunakan. Ukuran sampel ditentukan oleh peneliti didasarkan pada dua aspek yaitu aspek statistik dan aspek non statistik. Beberapa hal dalam aspek statistik untuk ukuran sampel adalah parameter yang akan ditaksir, tipe sampling yang digunakan tujuan penelitian, dan keragaman variabel yang diteliti. Sedangkan untuk aspek non statistik, ukuran sampel ditentukan oleh faktor waktu, biaya, dan ketersediaan satuan sampling.

Pemilihan teknik sampling didasarkan pada ukuran populasi, heterogenitas variabel yang diteliti, ketersediaan kerangka sampling dan dependensi antar unit. Sampling acak sederhana (Simple Random Sampling/SRS) mengharuskan setiap unit sampling dalam populasi memiliki peluang yang sama besar untuk terpilih ke dalam sampel (Cochran [2]). SRS dapat digunakan pada saat ukuran populasi berhingga, kerangka sampling tersedia dan antar unit sampling relatif homogen. Sampling acak berstrata (*Stratified Random Sampling*) digunakan pada saat kondisi populasi lebih heterogen karena apabila SRS digunakan pada kondisi populasi yang heterogen akan mengakibatkan presisi penaksir menjadi rendah. Pada proses sampling acak berstrata populasi dibagi ke dalam beberapa strata. Dengan stratifikasi, presisi suatu taksiran diharapkan menjadi semakin tinggi karenadidalam strata akan relative homogeny dibanding antar strata.Dengan kata lain bahwa variabilitas variabel yang akan diukur direfleksikan antar strata. Menurut Lohr [5] sampling klaster biasanya dilakukan apabila kerangka sampling yang memuat elemen atau unit observasi tidak tersedia. Pada sampling klaster unit sampling merupakan kumpulan atau kelompok (*cluster*) dari unit observasi. Masing-masing unit observasi di dalam populasi tepat berada dalam satu klaster yang merupakan unit sampling, makin banyak klaster yang dijadikan sampel maka variansnya akan lebih kecil oleh sebab itu varians dari penaksirnya tergantung pada keragaman antara rata-rata klaster. Untuk mendapatkan presisi yang terbaik, unit-unit

ISBN : 978.602.361.002.0

observasi di dalam masing-masing klaster harus heterogen dan rata-rata klaster harus serupa satu sama lain. Sampling klaster dapat dilakukan dua tahap yaitu pada tahap pertama memilih klaster yang disebut unit sampling primer dan pada tahap kedua memilih individu didalam klaster yang mana setiap pemilihan dilakukan secara SRS. Teknik ini disebut dengan *Two Stage Cluster Sampling*.Teknik ini dapat diperluas dengan mengkombinasikan berbagai teknik yang ada yang terdiri dari banyak tahap yang disebut Sampling Kompleks.

Data spasial adalah data dependen yang berasal dari suatu lokasi spasial yang berbeda dan mengindikasikan ketergantungan antara nilai pengukuran dengan lokasi (Cressie [3]). Untuk pemodelan untuk data spasial tidak dapat menggunakan model regresi klasik karena asumsi *error* saling bebas dan asumsi homogenitas tidak terpenuhi. Pendekatan analisis data spasial didasarkan atas tipe data spasial yaitu spatial titik atau spatial area. Pendekatan untuk jenis data spasial area diantaranya *Conditional Autoregressive Models* (CAR), *Mixed Regressive-Autoregressive* atau *Spatial Autoregressive Models* (SAR), *Spatial Error Models* (SEM), *Spatial Durbin Model* (SDM), dan *Spatial Autoregressive Moving Average* (SARMA). Dalam penelitian ini model yang digunakan adalah *Spatial Autoregressive Models* (SAR).

Asumsi *identic independen distribution* (i.i.d.) dari unit-unit sampling pada populasi spatial tidak dapat terpenuhi sehingga teknik sampling klasik tidak dapat digunakan. Pada dasarnya dalam sampling spasial mengadopsi adanya dependensi dan heterogenitas pada data. Pada prinsipnya secara geografis dapat diketahui bahwa pengamatan yang dekat satu sama lain lebih mungkin mirip (homogen) dan memiliki asosiasi yang kuat daripada pengamatan yang terpisah lebih jauhyang disebut hukum geografi Tobler. Beberapa hal yang harus diperhatikan dalam desain sampling untuk pemodelan spatial adalah pemilihan metode sampling dan perhitungan ukuran sampel, bagaimana cara pengambilan unit-unit sampling, *prior knowledge* (pengetahuan sebelumnya) untuk mendapatkan informasi yang dibutuhkan. Menurut Wang dkk., [8] teori sampling spasial dibagi menjadi dua yaitu sampling berbasis desain dan sampling berbasis model.

Pada sampling spasial sangat penting untuk mempertimbangkan adanya autokorelasi dan heterogenitas dari populasi yang akan disampel. Autokorelasi spasial melanggar asumsi independensi. Sementara heterogenitas dari bidang acak geografis terdiri atas *global variance* (varians antar wilayah) dan struktur spasial dari variasi tersebut (autokorelasi spasial populasi) (Wang dkk. [8]). Kedua unsur varians geografis tersebut perlu dipertimbangkan dalam merancang dan mengevaluasi desain sampel, termasuk rencana penentuan ukuran sampel dan penaksiran parameter (Cochran [3]).

2.1 MODEL *SPATIAL AUTOREGRESSIVE* (SAR)

Menurut LeSage [4]ada dua masalah yang muncul ketika data memiliki komponen wilayah yaitu adanya kebergantungan wilayah (*spatial dependence*) dan heterogenitas antar wilayah. Kebergantungan wilayahdiantara data pengamatan akan mengakibatkan terjadinya autokorelasi spasial (*spatial autocorrelation*) dan adanya heterogenitas spasial dalam model mengakibatkan varians yang tidak konstan. Kebergantungan wilayah dapat dianggap sebagai adanya hubungan fungsional antara apa yang terjadi pada satu titik dalam ruang dan apa yang terjadi di tempat lain. Menurut Tobler bahwa segala sesuatu saling berhubungan satu dengan yang lainnya, tetapi sesuatu yang lebih dekat mempunyai pengaruh daripada sesuatu yang jauh (Anselin [1]).

ISBN : 978.602.361.002.0

Model *Spatial Autoregressive Model*(SAR) yang disebut juga model spatial lag didasarkan atas parameter *autoregressive spatial*. Spasial *lag* terjadi akibat adanya ketergantungan nilai variabel respon suatu wilayah dengan nilai variabel respon wilayah lain. Dalam analisis data spatial dibutuhkan suatu ukuran untuk menyatakan kebergantungan antar satu wilyah dengan wilayah lainnya. Untuk itu dibentuklah matriks kebergantungan spasial (*contiguity*) yang disebut juga matriks pembobot spasial (*spatial weight*/ **W)**. Matriks **W**menggambarkan hubungan antar wilayah dan diperoleh berdasarkan informasi jarak atau ketetanggaan. Dalam penelitian ini matriks pembobot spasial yang digunakan adalah *Queen Contiguity*.Dalam kasus faktor-faktor yang memengaruhi tuberkulosis dalam Pontoh dkk [6] yang difokuskan di Kecamatan Ngamprah Kabupaten Bandung Barat matriks **W** dibentuk berdasarkan peta berikut ini:

Gambar 2.1 Peta Kecamatan Ngamprah

Matriks pembobot spatial tersebut distandardisasikan dengan persamaan berikut:

$$w_{ij} = \frac{c_{ij}}{\sum_{j=1}^{N} c_{ij}} \quad (2.1)$$

Model regresi spasial *lag* (*Spatial Autoregressive Model/SAR* Model) merupakan model regresi yang memperhatikan dependensi atau kebergantungan antar wilayah. Model umum persamaan regresi spasial *lag* dinyatakan pada persamaan berikut:

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.2)$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I}_n)$$

Dengan $\boldsymbol{y}$ adalah vector variable respon, $\boldsymbol{X}$ matriks variabel bebas, $\boldsymbol{\beta}$ koefisien regresi dan ρ adalah efek spatial berupa autokorelasi spatial. Parameter pada Persamaan 2 ditaksir menggunakan metode kemungkinan maksimum (*Maximum Likelihood Estimator/MLE*). (Anselin [1])

Fungsi fungsi *log-likelihood* seperti berikut untuk model 2 adalah:

$$L = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}ln\sigma^2 + ln|I - \rho W| - \frac{(y-\rho Wy-X\beta)'(y-\rho Wy-X\beta)}{2\sigma^2} \quad (2.3)$$

Dengan memaksimumkan fungsi log-likelihood tersebut maka diperoleh penaksir-penaksir parameter sebagai berikut:

ISBN : 978.602.361.002.0

$$\hat{\delta}^2 = \frac{(y-\rho Wy-X\hat{\beta})'(y-\rho Wy-X\hat{\beta})}{N} \quad (2.4)$$

$$\hat{\beta} = (X'X)^{-1}X'y - (X'X)^{-1}\hat{\rho}Wy \quad (2.5)$$

$$\hat{\rho} = (y'W'Wy)^{-1}y'W'y \quad (2.6)$$

2.2 UKURAN SAMPEL MODEL *SAR*

Secara umum ukuran sampel untuk data spatial didasarkan pada penambahan *spatial effect* pada rumus ukuran sampelnya. Vallejos dan Osorio [7] menjelaskan bahwa ukuran sampel untuk data spatial bias lebih efektif dibandingkan ukuran sampel menggunakan metode biasa. Vallejos dan Osorio [7] mendefinisikan ukuran sampel untuk data spatial sebagai *Effective Sample Size* (ESS). Rumus umum ESS adalah sebagai berikut

$$ESS = \mathbf{1}^t\mathbf{R}(\boldsymbol{\theta})^{-1}\mathbf{1} \quad (2.7)$$

Dengan $\mathbf{R}(\theta)$ adalah matriks autokorelasi dan θ adalah parameter.

Rumus pada Persamaan 2.7 berlaku untuk menaksir parameter rata-rata tunggal, parameter dua rata-rata, parameter model CAR dan parameter model SAR. Perbedaan untuk masing-masing penaksiran parameter adalah rumusan bagi $\mathbf{R}(\theta)$. Untuk model SAR rumus untuk invers $\mathbf{R}(\rho)$ adalah sebagai berikut

$$R_{SAR}^{-1} = C\Sigma^{-1}C \quad (2.8)$$

Dengan

$$\Sigma = (I-\rho W)^{-1}\Sigma_v(I-\rho W^t)^{-1} \quad (2.9)$$

$$C = diag(\sigma_1,\sigma_2,...,\sigma_n) \quad (2.10)$$

$$\sigma_i^2 = \Sigma_{ii} \quad (2.11)$$

Untuk $\mathbf{\Sigma_v} = \sigma^2\mathbf{I}$ dan maka ESS untuk model SAR adalah sebagai berikut

$$ESS_{SAR} = \mathbf{1}^t\mathbf{R}_{\mathbf{SAR}}^{-1}\mathbf{1} \quad (2.12)$$

3. HASIL PENELITIAN DAN PEMBAHASAN

Analisis yang digunakan dalam Pontoh dkk [6] mengenaipendeteksian faktor resiko serta pemetaan penyebaran tuberkolosis anak Kecamatan Ngamprah Kabupaten Bandung adalahmodel multilevel dengan efek spasial sebagai berikut

$$y = \beta_{0j} + X\beta + \varepsilon \quad (3.1)$$

$$\beta_{0j} = \rho W\beta_{0j} + v_j \quad (3.2)$$

Dengan

ISBN : 978.602.361.002.0

$$\varepsilon \sim N(0,\sigma^2)$$

$$v_j \sim N(0,\sigma_j^{\ 2})$$

Persamaan 3.1 merupakan model level 1 dan Persamaan 3.2 merupakan model level 2. Terlihat bahwa pada level 2 terdapat efek spatial antar wilayah.

Tujuan dari analisis tersebut adalah menaksir parameter pada model spatial yang digunakan yaitu model SAR sehingga desain sampling yang telah dijelaskan pada bagian sebelumnya dapat diterapkan untuk kasus ini. Informasi *Prior* untuk kasus ini diperoleh pada penelitian sebelumnya mengenai hal yang sama akan tetapi dengan tujuan yang berbeda. Tabel 1 menunjukkan distribusi unit sampling, dalam kasus ini RT, dari setiap desa di Kecamatan Ngamprah Kabupaten Bandung Barat dan nilai varians untuk setiap desa digunakan sebagai informasi prior. Dengan demikian penelitian selanjutnya terkendala oleh kurangnya informasi mengenai nilai autokorelasi (ρ). Informasi tersebut akan ditetapkan oleh peneliti dengan berbagai kemungkinan.

Tabel 3.1
Unit Sampling di Kecamatan Ngamprah Kabupaten Bandung Barat

No	Desa	Banyak RT	Varians
1	Bojong Koneng	73	3.7
2	Cilame	146	6.6
3	Cimanggu	40	3.3
4	Cimareme	42	2.1
5	Gadobangkong	64	3.9
6	Margajaya	73	5.6
7	Mekarsari	38	2.4
8	Ngamprah	34	1.4
9	Pakuhaji	40	3.1
10	Sukatani	34	2.4
11	Tanimulya	157	4.6
JUMLAH		**741**	

Sumber: BPS Jawa Barat

Menurut Vallejos dan Osorio [7] metode sampling yang digunakan menentukan perhitungan ukuran sampel. Dalam analisis data spatial banyak konfigurasi berbeda untuk wilayah sampling yang akan digunakan maka *stratified random sampling* merupakan metode yang cocok untuk digunakan. Dalam kasus ini konfigurasi yang digunakan didasarkan atas wilayah perbatasan desa sehingga *stratified random sampling* merupakan metode yang dipilih.

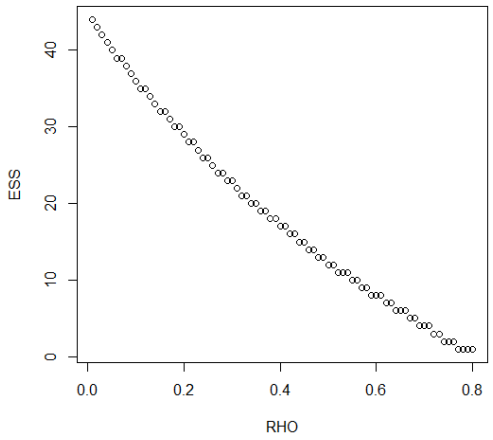
ISBN : 978.602.361.002.0

Untuk menentukan ukuran sampel perlu dibentuk terlebih dahulu matriks bobot spatial. Berdasarkan peta pada Gambar 1 dapat dibentuk matriks *Queen Continguity* sebagai berikut

$$W^* = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Selanjutnya matriks **W*** tersebut distandarisasi menggunakan Persamaan 1 menjadi matriks **W** yang akan digunakan dalam analisis.

Apabila desain sampling yang digunakan adalah stratifikasi maka ukuran sampel untuk kasus tersebut dengan distribusi unit sampling setiap kecamatan dan nilai varians setiap desa adalah seperti pada Tabel 1 adalah sebesar 44 RT. Apabila desain sampling stratifikasi tersebut melibatkan efek spatial yang ditunjukkan oleh matriks W, dengan beberapa nilai autokorelasi (ρ) maka ukuran sampel untuk kasus tersebut seperti terlihat pada gambar dibawah ini:



Gambar 3.1Nilai ESS terhadap autokorelasi (ρ)

Dari Gambar 2 tersebut jelas terlihat bahwa ukuran sampel sangat dipengaruhi oleh besarnya nilai autokorelasi spatial. Semakin besar nilai autokorelasi spatial maka nilai ESS semakin kecil. Artinya apabila antar desa memiliki nilai autokorelasi yang sangat besar maka ukuran sampel yang harus diambil hanya sedikit, sebaliknya dengan nilai autokorelasi spatial yang kecil maka ukuran sampel yang harus diambil semakin mendekati ukuran sampel

ISBN : 978.602.361.002.0

Stratifikasi. Pada saat ukuran sampel semakin kecil, bahkan mungkin lebih kecil dari banyak strata maka akan ada strata yang tidak memiliki wakil ke dalam sampel sehingga apabila desain stratifikasi tetap dipertahankan maka ukuran sampel minimal sama dengan banyak strata.

3. SIMPULAN

Desain sampling untuk data spatial dengan tujuan membuat pemodelan *spatial autoregresive* digunakan adalah sampling stratifikasi. Pada perhitungan ukuran sampel, efek spatial dalam hal ini autokorelasi spatial harus disertakan sehingga ESS merupakan metode yang tepat untuk digunakan.

Berdasarkan hasil perhitungan, untuk kasus apabila desain sampling yang digunakan adalah stratifikasi tanpa melibatkan efek spatial maka diperoleh ukuran sampel sebesar 44. Ukuran sampel ini akan semakin kecil apabila efek spatial dilibatkan dalam perhitungan menggunakan ESS yang diusulkan oleh Vallejos dan Osorio [7] seiring membesarnya nilai autokorelasi.

Dalam Pontoh dkk [5] distribusi dari model yang digunakan adalah distribusi poisson sedangkan dalam penelitian ini menggunakan distribusi normal, sehingga penelitian selanjutnya adalah membangun desain sampling pada pemodelan spatial dengan asumsi berdistibusi poisson.

DAFTAR PUSTAKA

[1] Anselin, L. (1988)*Spatial Econometrics : Methods and Models*. Dordrecht : Kluwer Academic Publisher

[2] Cochran, William G.(1977). *Sampling Techniques 3rd edition*. New York: John Wiley & Sons, Inc.

[3] Cressie, N., 1993. *Statistics for Spatial Data.* Wiley, New York

[4]Lesage, James P. (1998)*Spatial Econometrics.* Toledo: Department of Economics University of Toledo

[5] Lohr, Sharon. L. (1999). *Sampling Design and Analysis*. Duxbury Press.

[6] Pontoh, Resa S., Chadidjah, Anna, Faidah, Devi Yusti (2014) Pendeteksian Faktor Resiko serta Pemetaan Penyebaran Tuberkolosis Anak Kecamatan Ngamprah Kabupaten Bandung dengan Pendekatan Model Multilevel dengan Efek Spasial. Laporan Penelitian Departemen Statistika FMIPA UNPAD

[7] Vallejos, R., & Osorio, F. (2014). Effective Sample Size of Spatial Process Models. *Spatial Statistics.*

[8] Wang, J.-F., Stein, A., Gao, B.-B., & Ge, Y. (2012). A review of spatial sampling. *Spatial Statistics.*