

Case studies



Roberto Benedetti
University of Chieti-
Pescara, Italy



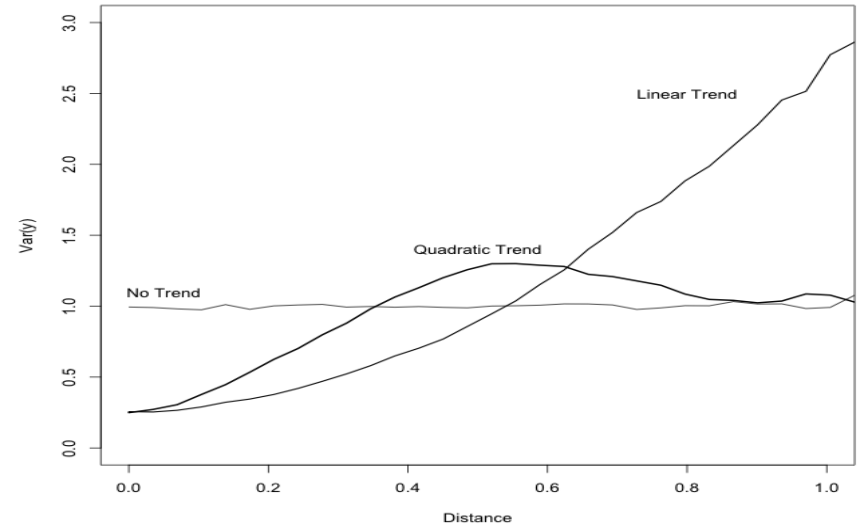
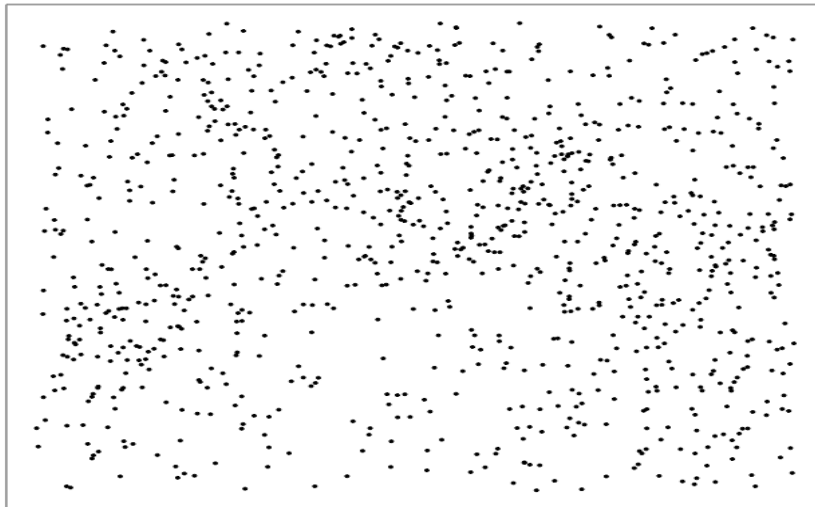
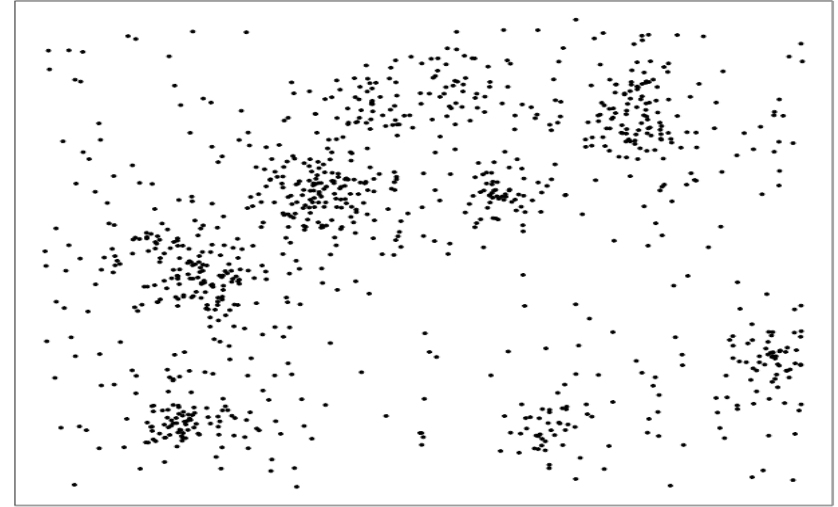
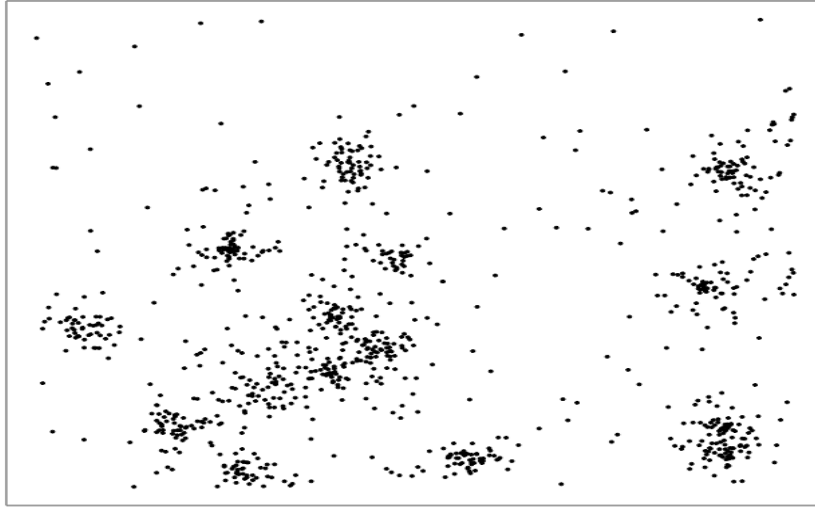
Federica Piersimoni
ISTAT, Italian
National Statistical
Institute, Rome,
Italy



Francesco Pantalone
University of Perugia,
Italy

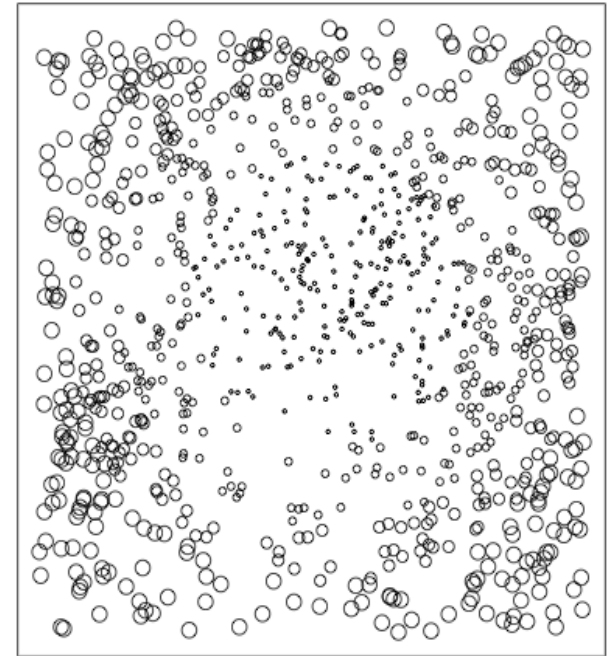
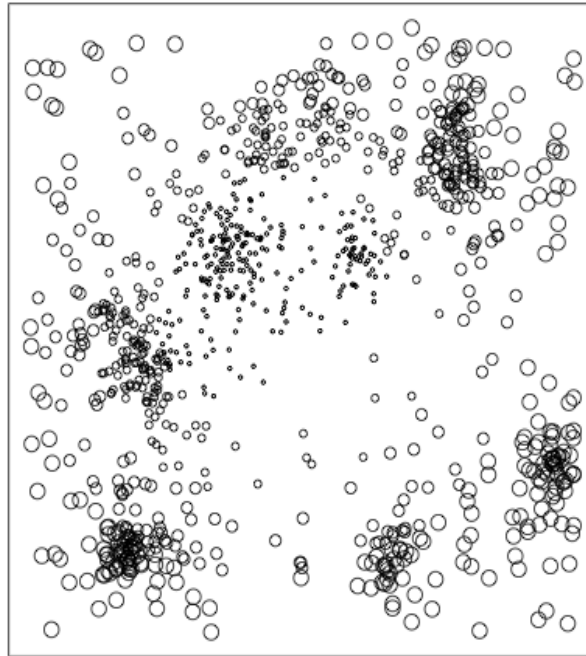
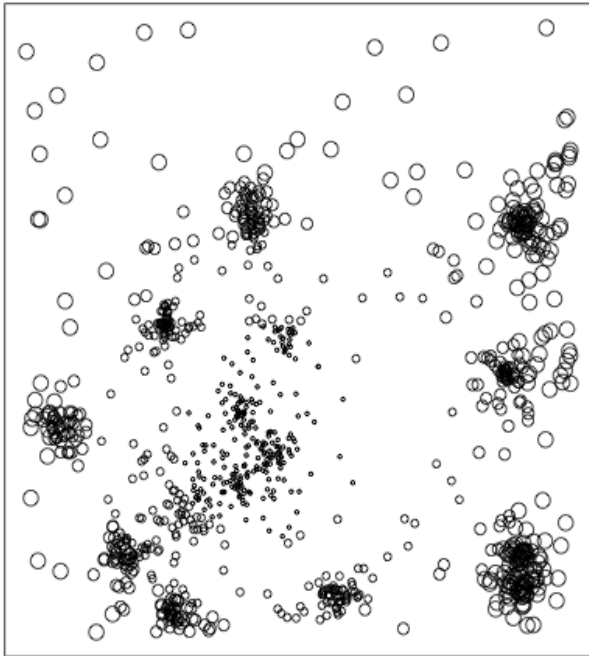


A Case Study: Simulated Populations



A Case Study: Simulated Populations

First order inclusion probabilities when using SWD without standerdization of the distance matrix



A Case Study: Simulated Populations

Design	n	No Trend		Linear Trend		Quadratic Trend		$V(v_k)$	Design	n	No Trend		Linear Trend		Quadratic Trend		$V(v_k)$
		Dependence		Dependence		Dependence					Dependence		Dependence		Dependence		
		Low	High	Low	High	Low	High				Low	High	Low	High	Low	High	
CUBE 1	10	1,00	0,99	0,56	0,48	0,89	0,94	0,33	CUBE 1	100	1,00	0,98	0,51	0,42	0,89	0,93	0,51
CUBE 2	10	1,00	0,95	0,57	0,50	0,55	0,53	0,27	CUBE 2	100	1,00	0,93	0,51	0,42	0,48	0,45	0,49
GRTS	10	1,00	0,80	0,68	0,58	0,67	0,66	0,21	GRTS	100	1,00	0,56	0,53	0,30	0,55	0,41	0,23
LPM 1	10	1,01	0,77	0,59	0,49	0,65	0,64	0,15	LPM 1	100	1,01	0,52	0,52	0,28	0,53	0,37	0,19
LPM 2	10	1,02	0,77	0,60	0,48	0,64	0,63	0,15	LPM 2	100	1,01	0,51	0,52	0,28	0,52	0,37	0,19
SCPS	10	1,01	0,78	0,62	0,50	0,66	0,66	0,17	SCPS	100	1,01	0,52	0,52	0,28	0,53	0,37	0,20
DBSS 1	10	1,00	0,99	0,57	0,48	0,90	0,93	0,33	DBSS 1	100	0,99	0,49	0,51	0,26	0,49	0,30	0,18
DBSS 2	10	1,00	0,95	0,58	0,49	0,55	0,53	0,27	DBSS 2	100	1,00	0,49	0,52	0,26	0,49	0,28	0,17
SWD 10	10	1,00	0,91	0,70	0,61	0,74	0,77	0,27	SWD 10	100	1,01	0,90	0,72	0,62	0,77	0,77	0,49
PWD 1	10	1,02	0,81	0,66	0,56	0,70	0,69	0,22	PWD 1	100	1,00	0,56	0,52	0,31	0,53	0,41	0,21
PWD 5	10	1,02	0,60	0,53	0,31	0,50	0,40	0,10	PWD 5	100	0,99	0,44	0,50	0,23	0,48	0,27	0,11
PWD 10	10	1,02	0,50	0,50	0,22	0,48	0,30	0,09	PWD 10	100	0,99	0,42	0,50	0,21	0,47	0,25	0,09
DBPWD 1,1	10	1,02	0,80	0,50	0,35	0,64	0,64	0,18	DBPWD 1,1	100	1,01	0,58	0,50	0,27	0,54	0,41	0,23
DBPWD 1,5	10	1,01	0,54	0,51	0,25	0,48	0,34	0,10	DBPWD 1,5	100	0,98	0,46	0,50	0,21	0,48	0,29	0,12
DBPWD 1,10	10	1,03	0,48	0,50	0,22	0,47	0,28	0,09	DBPWD 1,10	100	0,98	0,43	0,49	0,20	0,47	0,27	0,11
DBPWD 2,1	10	1,02	0,78	0,50	0,35	0,47	0,41	0,15	DBPWD 2,1	100	1,00	0,60	0,50	0,27	0,48	0,31	0,25
DBPWD 2,5	10	0,97	0,54	0,51	0,24	0,48	0,29	0,10	DBPWD 2,5	100	0,99	0,49	0,50	0,23	0,47	0,25	0,15
DBPWD 2,10	10	0,96	0,49	0,52	0,21	0,48	0,25	0,10	DBPWD 2,10	100	1,00	0,47	0,50	0,22	0,46	0,25	0,14

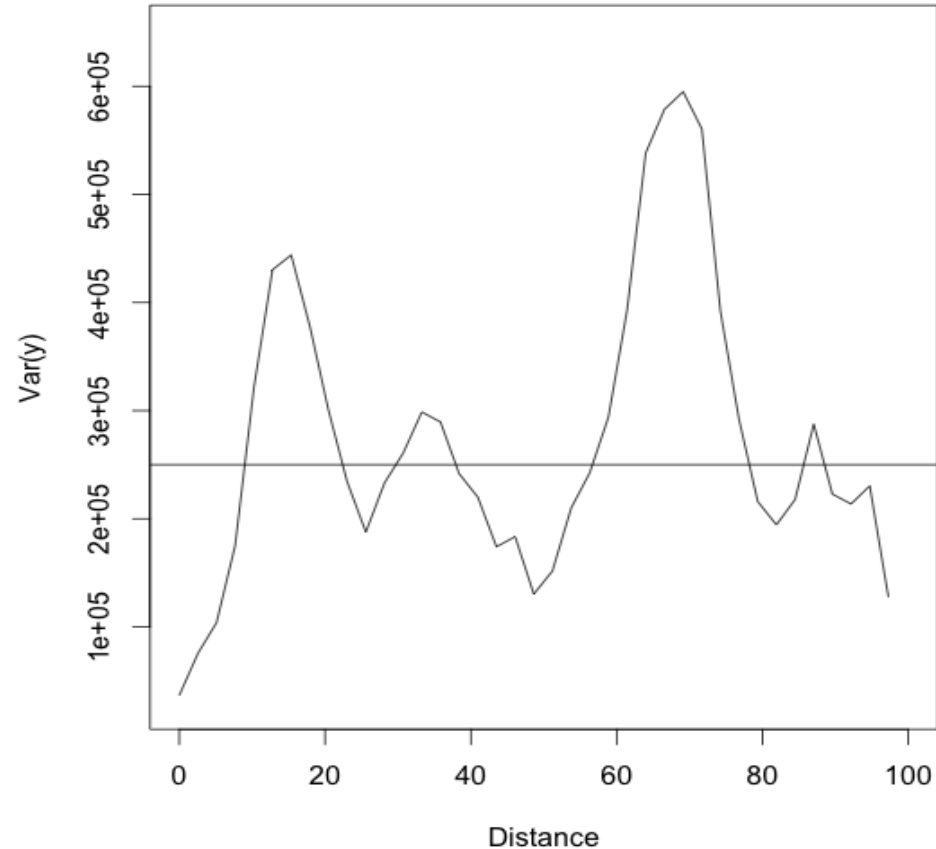
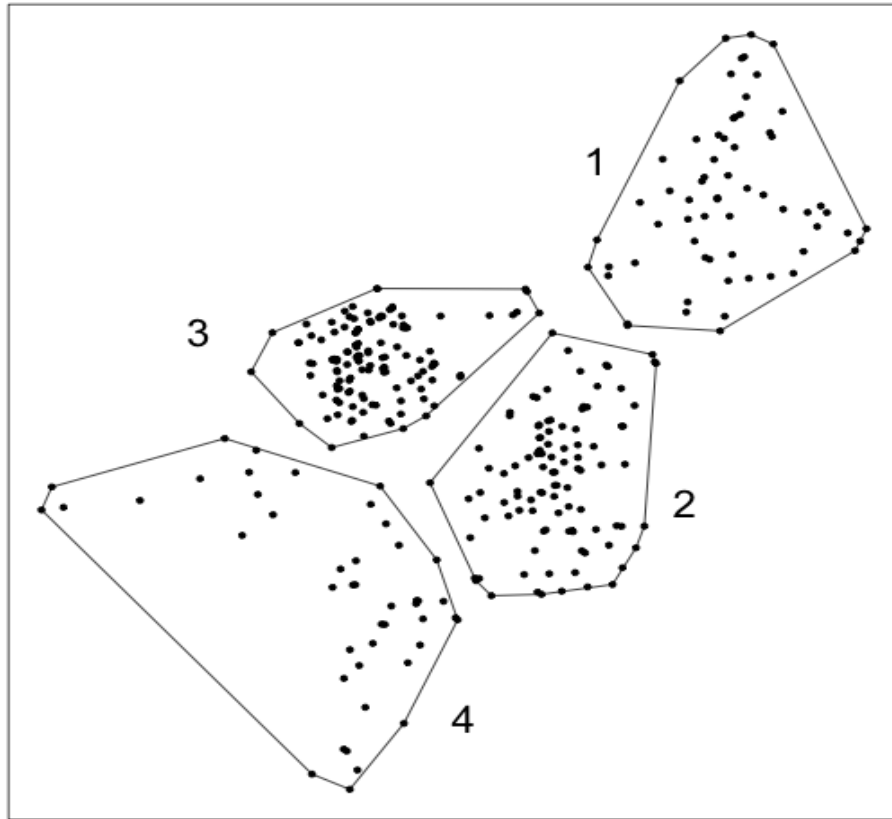
(MSE/MSE_{SRS}) and the mean of the spatial balance indices for each design estimated in 10,000 replicated samples in the highly clustered population for different sample sizes, trend, and homogeneity

A Case Study: Simulated Populations

Design	n	No Trend		Linear Trend		Quadratic Trend		$V(v_k)$	Design	n	No Trend		Linear Trend		Quadratic Trend		$V(v_k)$
		Dependence		Dependence		Dependence					Dependence		Dependence		Dependence		
		Low	High	Low	High	Low	High				Low	High	Low	High	Low	High	
CUBE 1	10	1,01	0,98	0,57	0,63	0,99	0,99	0,23	CUBE 1	100	0,99	0,99	0,51	0,60	1,01	1,00	0,33
CUBE 2	10	1,00	1,00	0,57	0,63	0,57	0,49	0,18	CUBE 2	100	1,00	0,98	0,51	0,58	0,51	0,41	0,31
GRTS	10	1,01	0,92	0,66	0,70	0,77	0,73	0,17	GRTS	100	1,01	0,67	0,53	0,42	0,60	0,45	0,14
LPM 1	10	1,00	0,90	0,61	0,63	0,77	0,72	0,12	LPM 1	100	1,02	0,63	0,52	0,38	0,55	0,38	0,10
LPM 2	10	1,01	0,89	0,63	0,63	0,76	0,72	0,12	LPM 2	100	1,03	0,63	0,52	0,39	0,56	0,38	0,11
SCPS	10	1,01	0,88	0,63	0,62	0,77	0,72	0,12	SCPS	100	1,01	0,63	0,51	0,39	0,55	0,38	0,10
DBSS 1	10	1,01	0,90	0,59	0,58	0,66	0,58	0,11	DBSS 1	100	1,02	0,60	0,51	0,37	0,52	0,31	0,10
DBSS 2	10	1,02	0,89	0,58	0,57	0,57	0,48	0,12	DBSS 2	100	1,01	0,58	0,51	0,36	0,51	0,29	0,10
SWD 10	10	1,01	0,96	0,71	0,71	0,80	0,77	0,20	SWD 10	100	1,00	0,97	0,72	0,73	0,83	0,79	0,31
PWD 1	10	1,01	0,93	0,68	0,68	0,78	0,75	0,17	PWD 1	100	1,01	0,70	0,53	0,45	0,56	0,43	0,15
PWD 5	10	1,00	0,89	0,56	0,58	0,64	0,57	0,09	PWD 5	100	1,01	0,54	0,49	0,35	0,51	0,29	0,08
PWD 10	10	0,99	0,87	0,53	0,55	0,60	0,50	0,06	PWD 10	100	1,01	0,52	0,49	0,34	0,51	0,27	0,07
DBPWD 1,1	10	1,00	0,94	0,50	0,53	0,81	0,76	0,13	DBPWD 1,1	100	1,00	0,70	0,50	0,41	0,58	0,44	0,15
DBPWD 1,5	10	1,00	0,88	0,51	0,50	0,66	0,57	0,08	DBPWD 1,5	100	1,02	0,56	0,49	0,34	0,51	0,30	0,08
DBPWD 1,10	10	1,00	0,86	0,50	0,48	0,61	0,50	0,06	DBPWD 1,10	100	1,00	0,52	0,48	0,34	0,51	0,28	0,07
DBPWD 2,1	10	1,01	0,94	0,50	0,54	0,52	0,41	0,11	DBPWD 2,1	100	1,02	0,73	0,50	0,42	0,50	0,31	0,16
DBPWD 2,5	10	0,99	0,87	0,50	0,50	0,51	0,38	0,07	DBPWD 2,5	100	1,00	0,56	0,49	0,36	0,49	0,26	0,09
DBPWD 2,10	10	0,99	0,86	0,50	0,48	0,51	0,36	0,05	DBPWD 2,10	100	1,01	0,52	0,48	0,35	0,49	0,25	0,08

(MSE/MSE_{SRs}) and the mean of the spatial balance indices for each design estimated in 10,000 replicated samples in the sparse population for different sample sizes, trend, and homogeneity

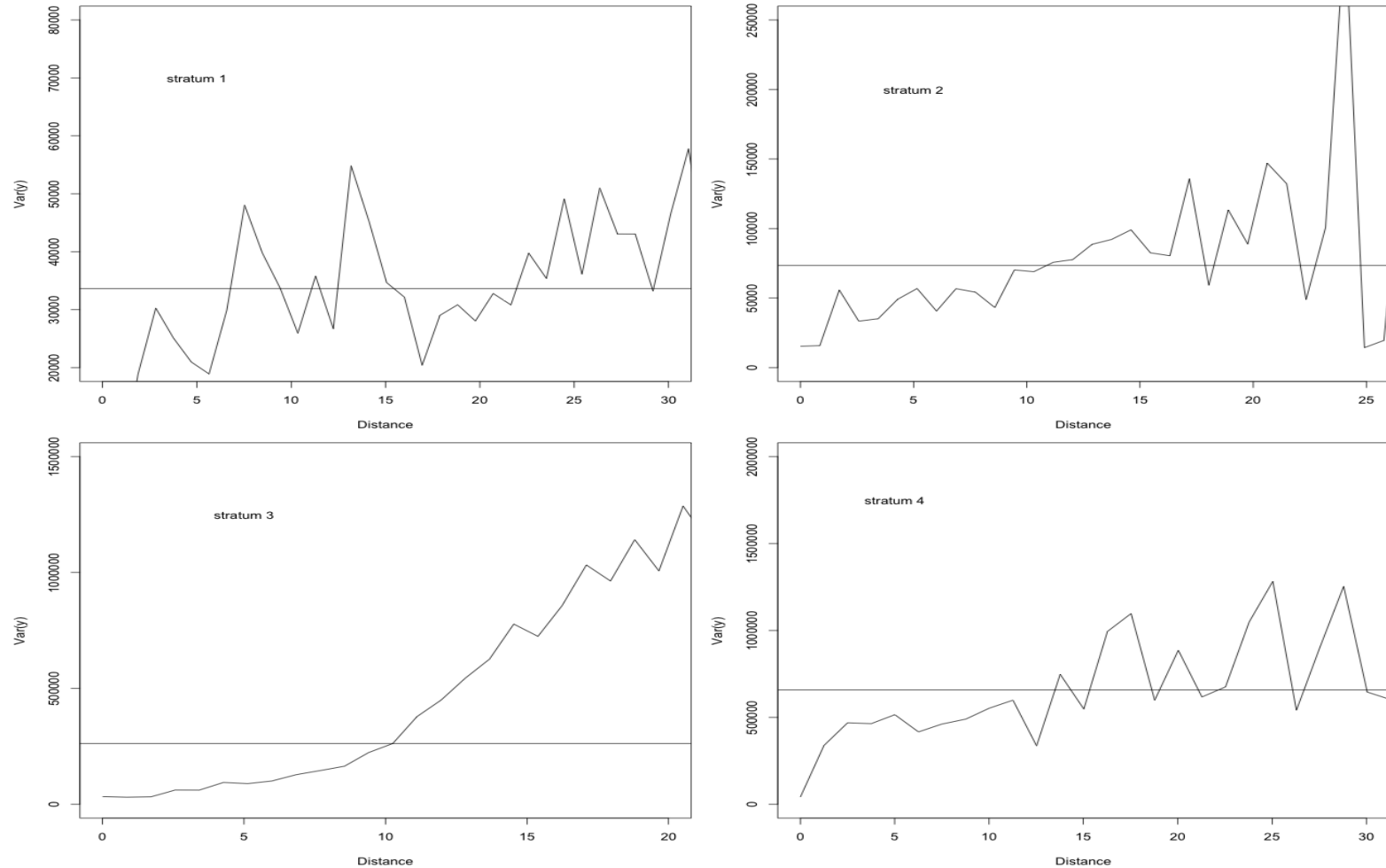
A Case Study: Assessing the Ecological Condition of Lakes in Northeastern USA



Spatial distribution of EMAP lakes population (left) (N=334) with the used partition in 4 strata and the variogram (right) of the “Acid Neutralising Capacity” (ANC) variable

A Case Study: Assessing the Ecological Condition of Lakes in Northeastern USA

Variograms of the ANC variable within each of the 4 strata

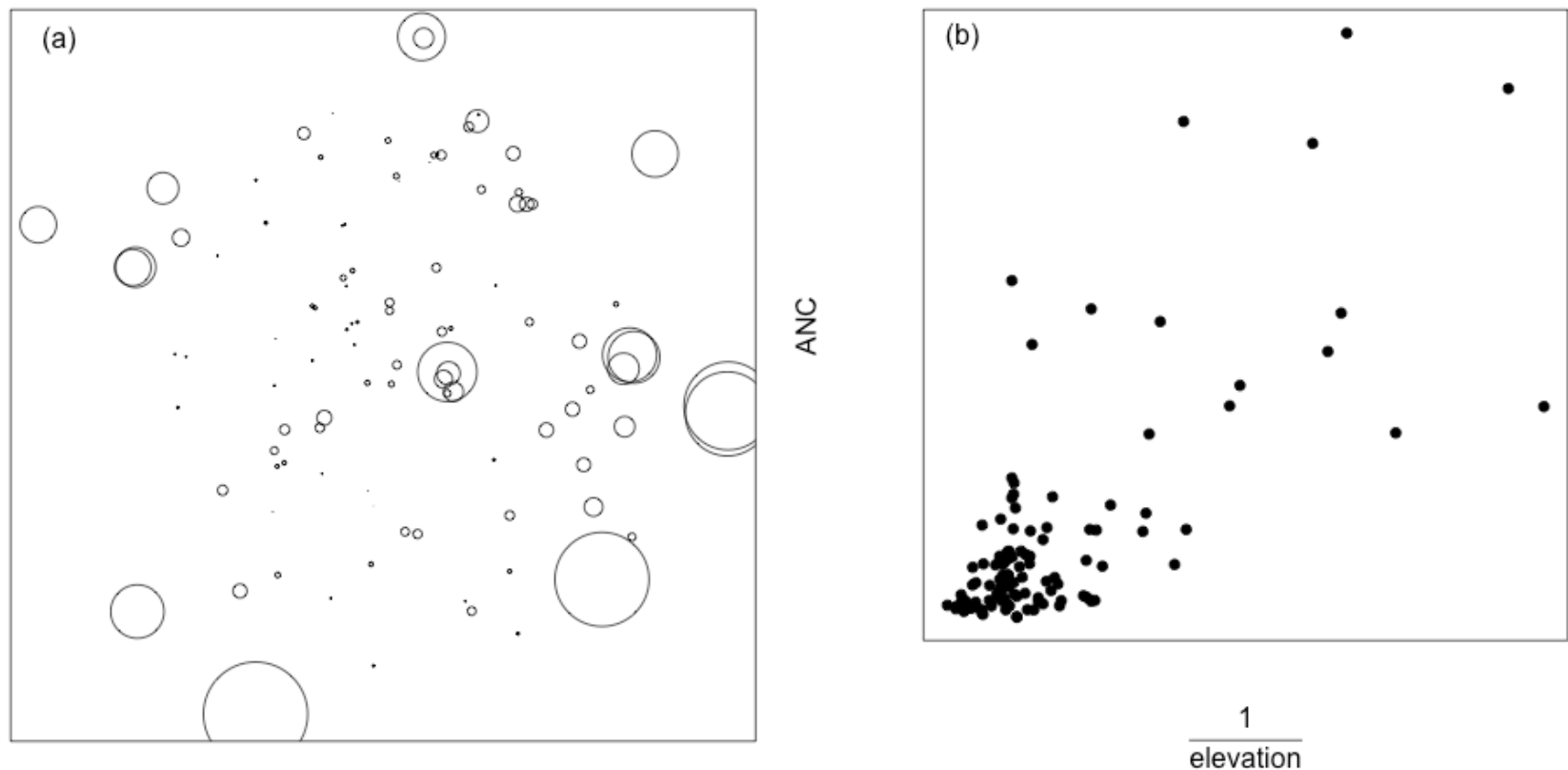


A Case Study: Assessing the Ecological Condition of Lakes in Northeastern USA

Design	Not Stratified			Stratified		
	$n=10$	$n=50$	$n=100$	$n=10$	$n=50$	$n=100$
SRS	1.00	1.00	1.00	0.52	0.77	0.82
GRTS	0.90	0.83	0.79	0.47	0.75	0.80
CUBE1	0.95	0.99	0.93	0.47	0.74	0.76
CUBE2	0.95	0.97	0.91	0.45	0.61	0.62
SCPS	0.80	0.66	0.56	0.44	0.56	0.53
LPM1	0.75	0.62	0.46	0.42	0.47	0.45
LPM2	0.81	0.67	0.54	0.48	0.58	0.54
PWD1	0.80	0.66	0.56	0.44	0.56	0.53
PWD5	0.72	0.56	0.42	0.39	0.44	0.39
PWD10	0.73	0.59	0.41	0.41	0.38	0.38

A Case Study: Ecological Condition of Lakes

Spatial distribution of EMAP lakes population, the size of the circles are proportional to the ANC variable. (b) Scatterplot of ANC and $1/\text{elevation}$.

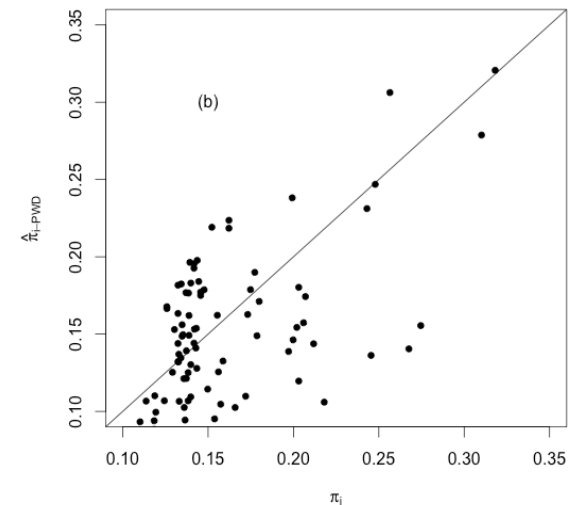
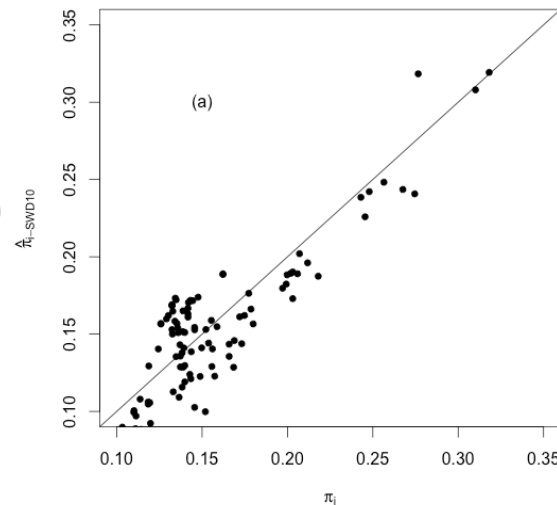


A Case Study: Assessing the Ecological Condition of Lakes in Northeastern USA

Relative efficiency of the sample mean (MSE/MSESRS) of the EMAP data-set estimated in 100,000 replicated samples for different designs and for the PWD and the SWD for not standardized and standardized distance matrix for constant or variable π_i -es

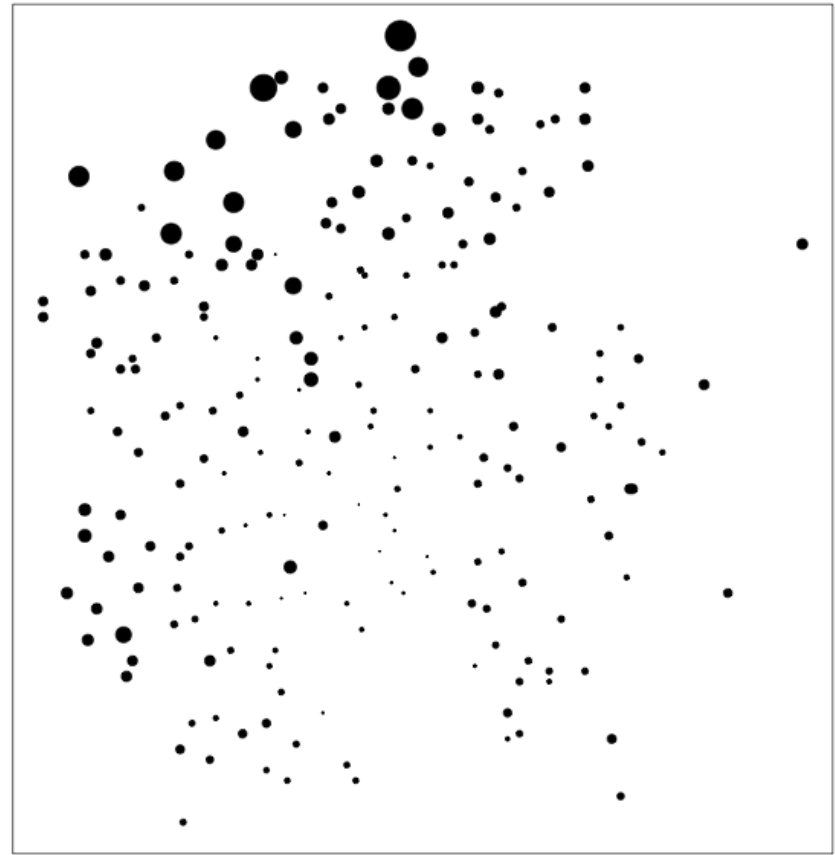
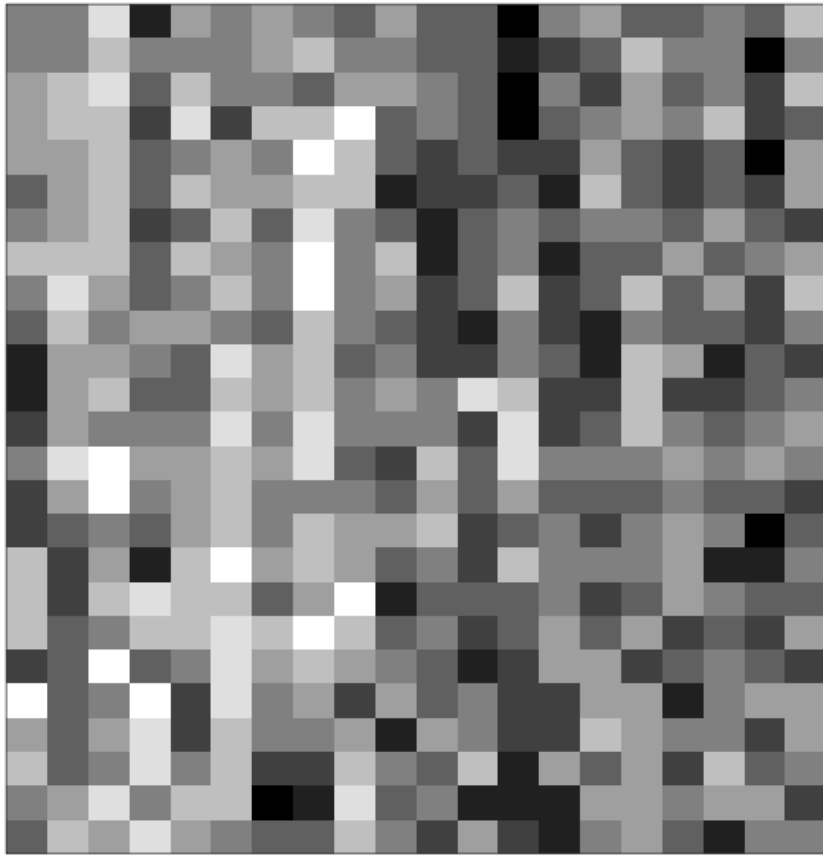
D	SWD						GRTS	SCPS	LPM1	LPM2
	PWD	b=10	b=20	b=30	b=40	b=50				
NotSTD	0.696	0.725	1.699	2.488	2.809	2.837	1.060	0.758	0.673	0.718
STD	0.724	0.861	0.803	0.775	0.751	0.740	-----	-----	-----	-----
pps	0.592	0.582	0.576	0.579	0.572	0.578	0.822	0.616	0.585	0.589

Scatterplot of the fixed and estimated first order inclusion probabilities for the (a) SWD10 and (b) PWD



A Case Study: Mercer-Hall Wheat Yield and Baltimore House Prices

Spatial distribution of the variable grain in Mercer-Hall data-set (left) (N=400) and of the variable price in Baltimore data-set (right) (N=211)

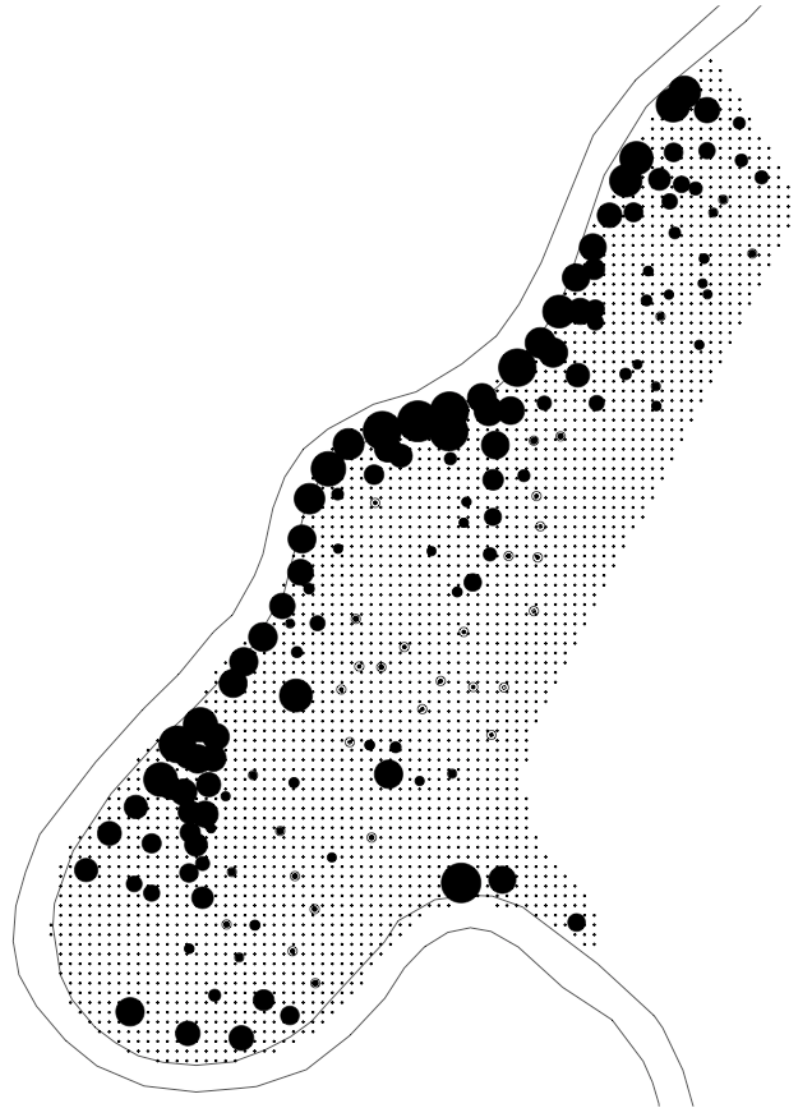


A Case Study: Mercer-Hall Wheat Yield and Baltimore House Prices

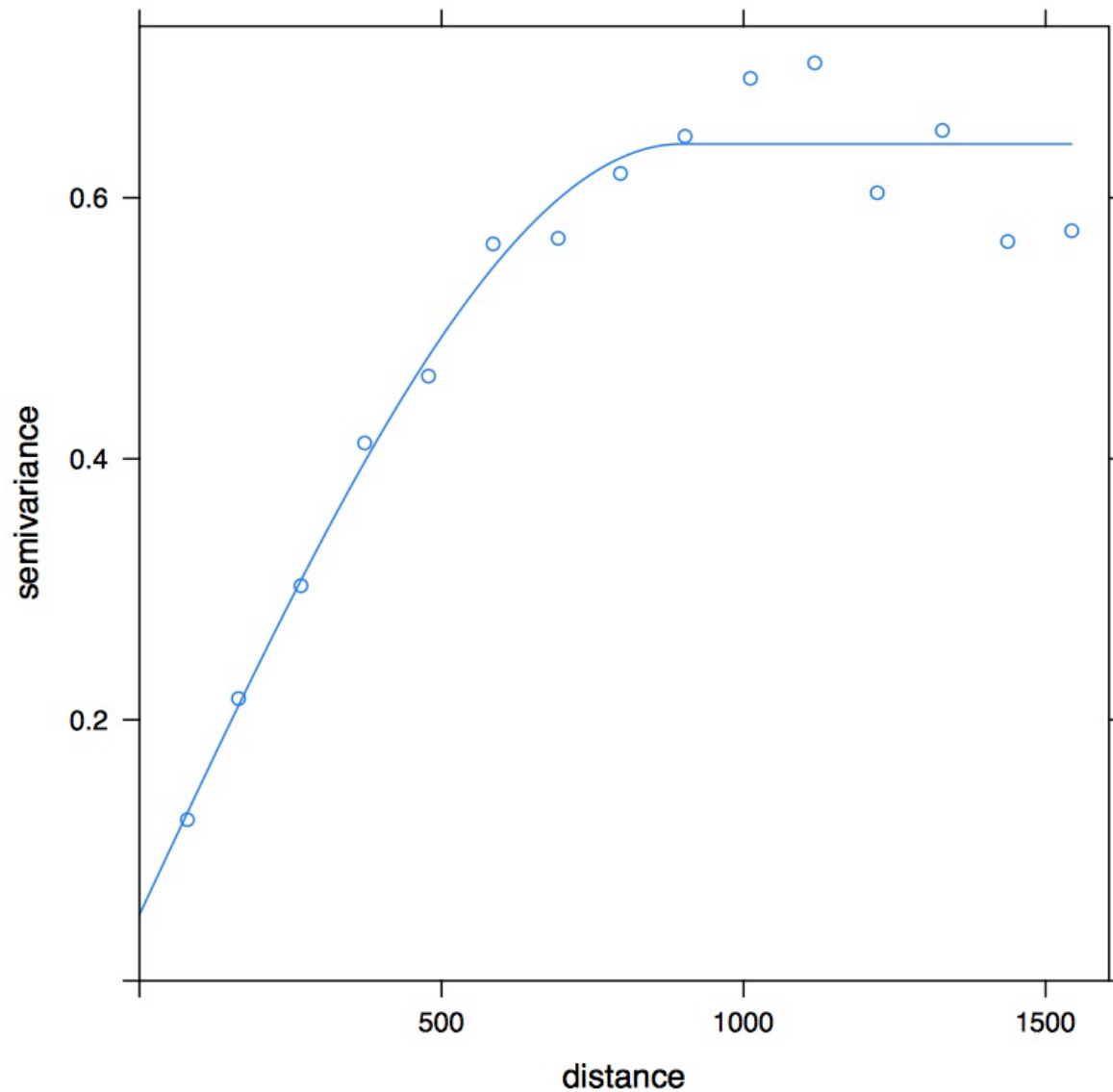
Design	Mercer-Hall						Baltimore					
	n	MSE/MSE_{SRS}	$V(v_k)$	n	MSE/MSE_{SRS}	$V(v_k)$	n	MSE/MSE_{SRS}	$V(v_k)$	n	MSE/MSE_{SRS}	$V(v_k)$
GRTS	10	0.54	0.14	100	0.23	0.12	10	0.58	0.16	50	0.35	0.17
CUBE 1	10	0.29	0.20	100	0.10	0.23	10	0.30	0.22	50	0.15	0.29
CUBE 2	10	0.32	0.16	100	0.11	0.21	10	0.35	0.17	50	0.18	0.26
DUST 1	10	0.57	0.22	100	0.34	0.87	10	0.80	0.30	50	2.86	0.89
DUST 2	10	0.59	0.21	100	0.34	0.90	10	0.76	0.28	50	2.80	0.97
SCPS	10	0.39	0.08	100	0.10	0.10	10	0.38	0.10	50	0.21	0.14
LPM 1	10	0.44	0.10	100	0.18	0.10	10	0.45	0.12	50	0.27	0.14
LPM 2	10	0.44	0.10	100	0.18	0.10	10	0.45	0.12	50	0.27	0.14
DBSS 1	10	0.34	0.09	100	0.12	0.10	10	0.36	0.11	50	0.21	0.14
DBSS 2	10	0.34	0.09	100	0.12	0.09	10	0.38	0.11	50	0.22	0.14
PWD 1	10	0,54	0,15	100	0,22	0,12	10	0,55	0,17	50	0,33	0,17
PWD 5	10	0,30	0,07	100	0,11	0,10	10	0,29	0,08	50	0,18	0,12
PWD 10	10	0,22	0,05	100	0,09	0,10	10	0,20	0,05	50	0,16	0,11
DBPWD 1,1	10	0,30	0,11	100	0,13	0,12	10	0,40	0,13	50	0,09	0,18
DBPWD 1,5	10	0,28	0,06	100	0,09	0,10	10	0,21	0,07	50	0,07	0,13
DBPWD 1,10	10	0,27	0,05	100	0,06	0,10	10	0,22	0,05	50	0,08	0,12
DBPWD 2,1	10	0,35	0,09	100	0,14	0,12	10	0,52	0,11	50	0,12	0,18
DBPWD 2,5	10	0,32	0,05	100	0,10	0,10	10	0,42	0,07	50	0,11	0,13
DBPWD 2,10	10	0,31	0,04	100	0,06	0,10	10	0,43	0,05	50	0,15	0,12

A Case Study: Meuse data set

The meuse data set is a data set comprising of four heavy metals measured in the top soil in a flood plain along the river Meuse. The governing process seems that polluted sediment is carried by the river, and mostly deposited close to the river bank. (The meuse data set: a brief tutorial for the gstat R package, Edzer Pebesma April 2, 2014; Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. Computers & Geosciences 30: 683-691) N = 155



A Case Study: Meuse data set



A Case Study: Meuse data set

Design	n	x	y	x^2	y^2	<i>cadmium</i>	<i>copper</i>	<i>lead</i>	<i>zinc</i>	$V(v_k)$	n	x	y	x^2	y^2	<i>cadmium</i>	<i>copper</i>	<i>lead</i>	<i>zinc</i>	$V(v_k)$
SRS	10	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,42	30	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	0,36
GRTS	10	0,29	0,30	0,29	0,30	0,92	0,88	0,92	0,93	0,29	30	0,18	0,19	0,18	0,19	0,92	0,88	0,92	0,92	0,31
CUBE 1	10	0,34	0,40	0,34	0,40	0,89	0,85	0,92	0,91	0,21	30	0,21	0,25	0,21	0,25	0,90	0,86	0,92	0,91	0,27
CUBE 2	10	0,44	0,44	0,44	0,44	0,90	0,89	0,88	0,91	0,17	30	0,31	0,29	0,31	0,29	0,88	0,85	0,84	0,88	0,16
SCPS	10	0,39	0,37	0,39	0,37	0,88	0,87	0,86	0,90	0,13	30	0,26	0,25	0,26	0,25	0,81	0,77	0,75	0,80	0,12
LPM 1	10	0,38	0,35	0,38	0,35	0,89	0,87	0,87	0,89	0,13	30	0,25	0,23	0,25	0,23	0,81	0,78	0,76	0,80	0,12
LPM 2	10	0,35	0,32	0,35	0,32	0,88	0,86	0,85	0,89	0,12	30	0,24	0,21	0,24	0,21	0,82	0,78	0,76	0,79	0,12
DBSS 1	10	0,32	0,34	0,32	0,34	0,80	0,77	0,80	0,82	0,13	30	0,20	0,22	0,20	0,22	0,72	0,67	0,70	0,71	0,12
DBSS 2	10	0,38	0,45	0,38	0,45	0,82	0,78	0,78	0,82	0,13	30	0,24	0,30	0,24	0,30	0,76	0,70	0,71	0,73	0,12
PWD 1	10	0,45	0,44	0,45	0,44	0,91	0,89	0,91	0,92	0,18	30	0,31	0,30	0,31	0,30	0,84	0,81	0,81	0,83	0,17
PWD 5	10	0,24	0,23	0,24	0,23	0,77	0,78	0,77	0,80	0,09	30	0,16	0,16	0,16	0,16	0,76	0,70	0,72	0,74	0,10
PWD 10	10	0,17	0,16	0,17	0,16	0,70	0,75	0,73	0,75	0,06	30	0,14	0,14	0,14	0,14	0,74	0,68	0,69	0,71	0,08
DBPWD 1,1	10	0,02	0,01	0,02	0,01	0,88	0,85	0,88	0,90	0,17	30	0,01	0,01	0,01	0,01	0,85	0,81	0,82	0,84	0,18
DBPWD 1,5	10	0,02	0,01	0,02	0,01	0,79	0,80	0,80	0,83	0,09	30	0,02	0,02	0,02	0,02	0,79	0,72	0,74	0,77	0,11
DBPWD 1,10	10	0,03	0,02	0,03	0,02	0,73	0,77	0,75	0,78	0,07	30	0,07	0,07	0,07	0,07	0,77	0,71	0,74	0,75	0,09
DBPWD 2,1	10	0,03	0,01	0,03	0,01	0,89	0,85	0,88	0,91	0,17	30	0,02	0,01	0,02	0,01	0,85	0,80	0,82	0,85	0,18
DBPWD 2,5	10	0,03	0,01	0,03	0,01	0,79	0,78	0,79	0,82	0,10	30	0,02	0,01	0,02	0,01	0,78	0,73	0,75	0,77	0,12
DBPWD 2,10	10	0,03	0,02	0,03	0,02	0,73	0,76	0,74	0,77	0,07	30	0,07	0,07	0,07	0,07	0,77	0,71	0,73	0,75	0,10

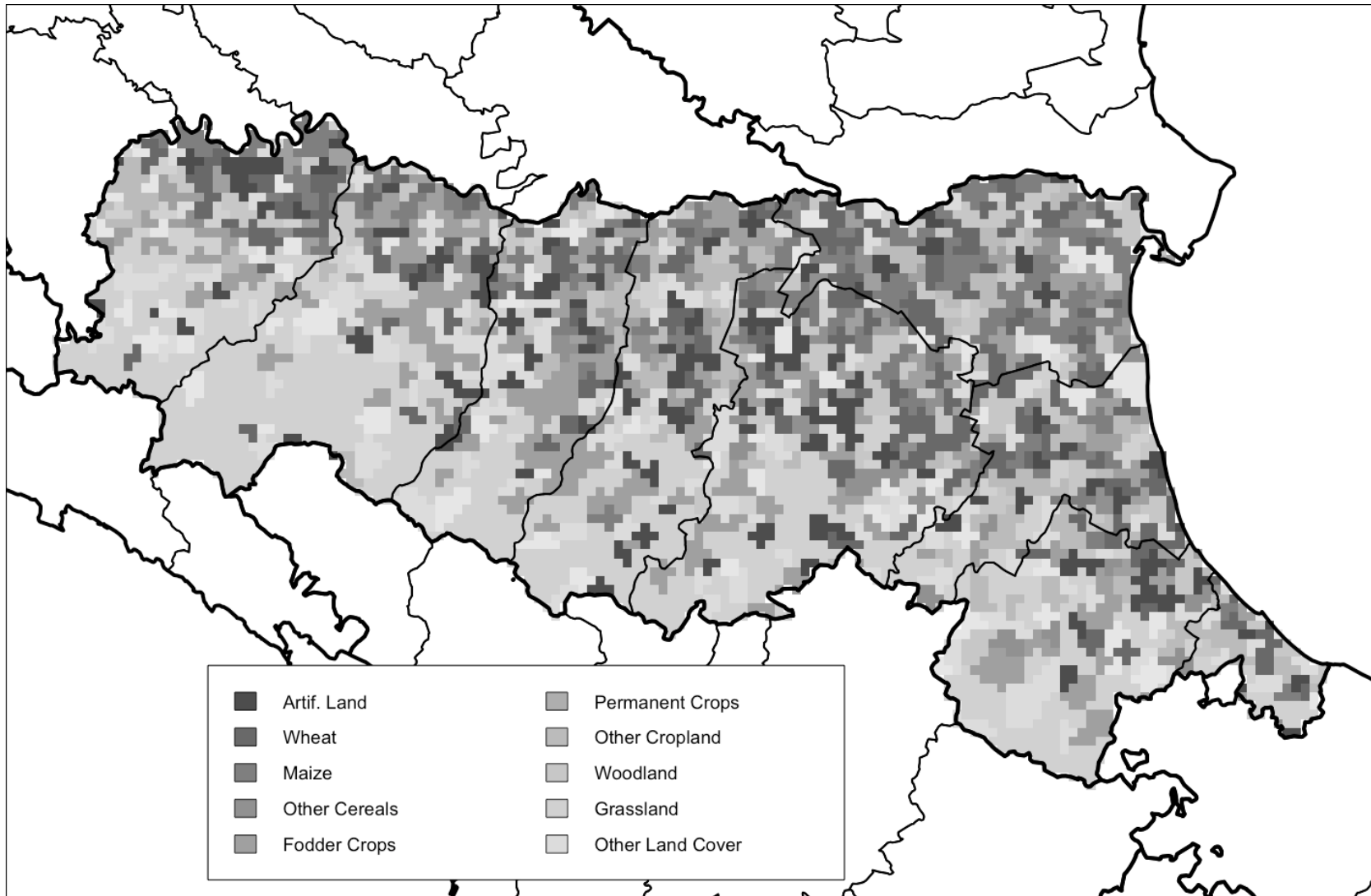
A Case Study: LUCAS Emilia Romagna

LUCAS survey Emilia Romagna Land Land Cover Estimates 2012. N = 5527



A Case Study: LUCAS Emilia Romagna

LUCAS survey Emilia Romagna Land Land Cover Estimates 2012. N = 5527



A Case Study: LUCAS Emilia Romagna

Area		n	Land Cover									Land Use					V(v _k)	
			Artif.	Wheat	Maize	Other	Fodder	Perman.	Other	Woodland	Grassland	Other Land	Agric.	Forestry	Urban	Unused		Other
			Land			Cereals	Crops	Crops	Cropland			Cover						Land Use
			1688,00	2380,00	1764,00	692,00	3120,00	1452,00	1368,00	5976,00	2368,00	1300,00	12884,00	4884,00	2180,00	1744,00	416,00	
CUBE 1	100		1,00	0,98	0,97	1,01	0,99	0,99	0,97	0,87	0,99	1,01	0,89	0,86	1,00	0,99	1,01	0,33
CUBE 2	100		0,99	0,97	0,95	1,00	0,99	0,98	0,99	0,88	0,98	1,01	0,90	0,86	1,00	0,99	0,99	0,31
GRTS	100		0,98	0,95	0,93	0,98	0,94	0,94	0,96	0,82	0,97	0,98	0,86	0,77	0,96	0,94	0,94	0,29
LPM1	100		0,96	0,95	0,94	0,97	0,94	0,90	0,95	0,80	0,95	0,94	0,85	0,75	0,94	0,92	0,92	0,11
LPM2	100		0,97	0,94	0,93	0,97	0,94	0,91	0,94	0,81	0,96	0,96	0,85	0,77	0,94	0,93	0,92	0,07
SCPS	100		0,96	0,94	0,95	0,97	0,92	0,92	0,94	0,80	0,95	0,96	0,84	0,75	0,93	0,93	0,92	0,07
PWD 1	100		0,97	0,95	0,94	0,96	0,94	0,94	0,96	0,82	0,96	0,95	0,88	0,78	0,96	0,94	0,94	0,05
PWD 5	100		0,96	0,94	0,92	0,97	0,92	0,91	0,94	0,80	0,95	0,95	0,85	0,75	0,93	0,93	0,93	0,08
PWD 10	100		0,95	0,93	0,94	0,96	0,92	0,91	0,95	0,79	0,96	0,93	0,83	0,74	0,94	0,90	0,94	0,04
CUBE 1	600		0,99	0,97	0,95	1,00	1,00	0,98	0,97	0,87	1,00	1,00	0,89	0,86	1,00	0,96	1,01	0,31
CUBE 2	600		0,99	0,97	0,95	1,00	1,00	0,98	0,98	0,87	1,00	1,00	0,89	0,85	1,00	0,95	1,00	0,30
GRTS	600		0,90	0,87	0,85	0,90	0,86	0,84	0,87	0,73	0,90	0,89	0,78	0,70	0,88	0,83	0,82	0,30
LPM1	600		0,85	0,84	0,81	0,84	0,82	0,78	0,82	0,68	0,84	0,84	0,73	0,65	0,84	0,77	0,76	0,11
LPM2	600		0,85	0,83	0,82	0,84	0,82	0,79	0,83	0,68	0,86	0,84	0,73	0,65	0,82	0,78	0,78	0,07
SCPS	600		0,85	0,82	0,81	0,82	0,81	0,77	0,82	0,68	0,84	0,83	0,72	0,65	0,83	0,76	0,75	0,08
PWD 1	600		0,91	0,88	0,86	0,90	0,87	0,85	0,88	0,73	0,91	0,90	0,78	0,70	0,89	0,84	0,85	0,07
PWD 5	600		0,85	0,84	0,79	0,82	0,81	0,77	0,81	0,67	0,84	0,82	0,72	0,65	0,82	0,76	0,76	0,08
PWD 10	600		0,84	0,82	0,79	0,82	0,81	0,76	0,80	0,66	0,83	0,82	0,71	0,64	0,82	0,76	0,74	0,06

A Case Study: LUCAS Emilia Romagna

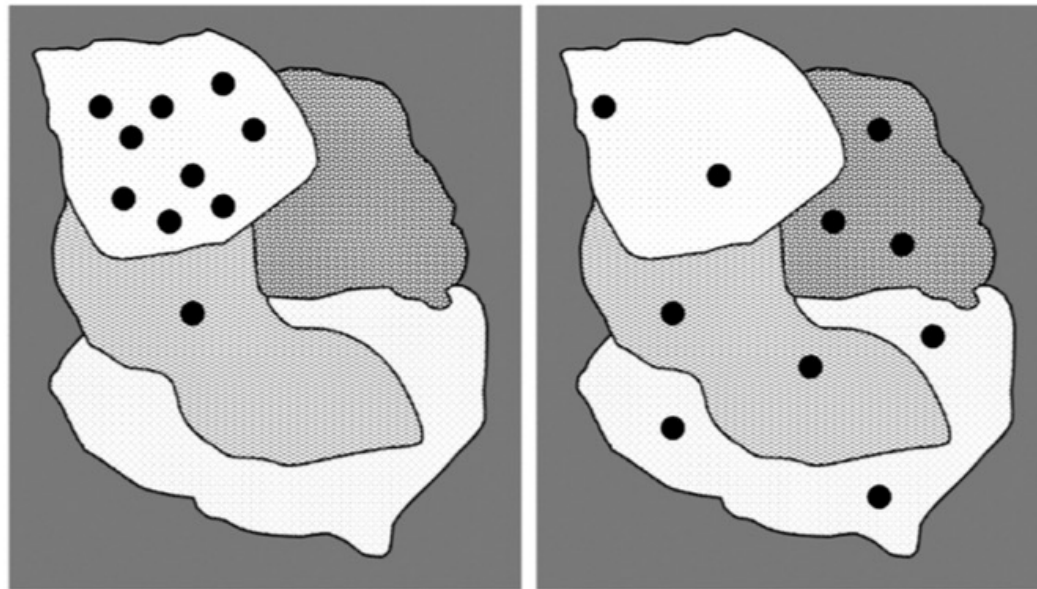
	<i>n</i>	<i>Piacenza</i>	<i>Parma</i>	<i>R. Nell'Emilia</i>	<i>Modena</i>	<i>Bologna</i>	<i>Ferrara</i>	<i>Ravenna</i>	<i>Forlì Cesena</i>	<i>Rimini</i>	<i>% Missing</i>
<i>Area</i>		2592	3460	2276	2696	3692	2628	1868	2364	532	<i>Values</i>
SRS	100	0,27	0,23	0,29	0,27	0,22	0,27	0,33	0,29	0,63	8,69
CUBE 1	100	0,23	0,20	0,29	0,27	0,22	0,21	0,31	0,23	0,60	7,50
CUBE 2	100	0,17	0,20	0,27	0,24	0,20	0,21	0,29	0,20	0,56	6,17
GRTS	100	0,11	0,10	0,16	0,15	0,11	0,13	0,19	0,13	0,32	0,24
LPM1	100	0,08	0,08	0,13	0,12	0,09	0,10	0,16	0,10	0,27	0,00
LPM2	100	0,09	0,09	0,13	0,12	0,09	0,10	0,16	0,10	0,26	0,00
SCPS	100	0,07	0,07	0,11	0,11	0,08	0,09	0,15	0,11	0,25	0,00
PWD 1	100	0,10	0,10	0,16	0,14	0,10	0,12	0,19	0,12	0,32	0,27
PWD 5	100	0,07	0,07	0,12	0,11	0,08	0,09	0,14	0,09	0,22	0,00
PWD 10	100	0,06	0,07	0,10	0,10	0,07	0,08	0,13	0,08	0,21	0,00
SRS	600	0,11	0,09	0,11	0,10	0,09	0,11	0,13	0,11	0,24	0,00
CUBE 1	600	0,09	0,08	0,11	0,10	0,08	0,08	0,12	0,09	0,23	0,00
CUBE 2	600	0,07	0,08	0,11	0,09	0,08	0,08	0,11	0,08	0,21	0,00
GRTS	600	0,03	0,03	0,04	0,04	0,03	0,03	0,05	0,03	0,08	0,00
LPM1	600	0,02	0,02	0,03	0,03	0,02	0,03	0,04	0,03	0,07	0,00
LPM2	600	0,02	0,02	0,03	0,03	0,03	0,03	0,04	0,03	0,07	0,00
SCPS	600	0,02	0,02	0,03	0,03	0,02	0,02	0,04	0,03	0,06	0,00
PWD 1	600	0,02	0,03	0,04	0,04	0,03	0,03	0,05	0,03	0,08	0,00
PWD 5	600	0,02	0,02	0,03	0,03	0,02	0,02	0,04	0,02	0,06	0,00
PWD 10	600	0,02	0,02	0,03	0,03	0,02	0,02	0,04	0,02	0,06	0,00

A Case Study: Sample Design for Small Area Estimation

Spatially balanced samples should reduce the variability of the sample size in each not planned domain

$$\tilde{t}_d = N_d \frac{\sum_{s_d} (y_k / \pi_k)}{\sum_{s_d} (1 / \pi_k)}$$

A design cannot modify the goodness of fit of a model but can reduce the direct estimator variance.



A Case Study: Sample Design for Small Area Estimation

Spatially balanced samples as they should reduce the variability of the sample size in each not planned domain

$$\tilde{t}_d = N_d \frac{\sum_{s_d} (y_k / \pi_k)}{\sum_{s_d} (1 / \pi_k)} \quad \hat{V}(\tilde{t}_d) = \left(\frac{N_d}{\hat{N}_d} \right)^2 \sum \sum_{s_d} \tilde{\Delta}_{kl} \left(\frac{y_k - \tilde{y}_{s_d}}{\pi_k} \right) \left(\frac{y_l - \tilde{y}_{s_d}}{\pi_l} \right) \quad \hat{N}_d = \sum_{s_d} (1 / \pi_k)$$

When and there is sufficient sample observations available for d-th SA, the HT direct estimator is unbiased, with respect to the design, but not efficient. In SA estimation problems, with inadequate sample, this estimator can be biased and unreliable. Furthermore, it is characterized by very high variance. As a consequence, it is not often used in practical applications. However a design cannot modify the goodness of fit of a model but can reduce the direct estimator variance.

A Case Study: Sample Design for Small Area Estimation

Fay – Harriot Model – Area level Approach

- Empirical BLUP (EBLUP) for the parameter of interest θ

$$\tilde{\theta}_i = \hat{\Phi}_i \hat{\theta}_i + (1 - \hat{\Phi}_i)(\mathbf{x}'_i \hat{\boldsymbol{\beta}} + \mathbf{z}'_i \hat{\mathbf{u}})$$

with $\hat{\Phi}_i = \frac{z_i^2 \hat{\sigma}_u^2}{z_i^2 \hat{\sigma}_u^2 + \sigma_e^2}$ *shrinkage factor*.

The shrinkage factor is used to assign the weights to the estimators according to the number of observations of the domain: the higher it is, the more the weight of the direct estimator will be. At the opposite, the lower the number of observations, the higher the weight of the model estimator will be.

A Case Study: Sample Design for Small Area Estimation

$$MSE\left(\tilde{\theta}_d^{EBLUP}\right) \approx g_{1d}\left(\sigma_v^2\right)+g_{2d}\left(\sigma_v^2\right)+g_{3d}\left(\sigma_v^2\right)$$

$$g_{1d}\left(\sigma_v^2\right)=\gamma_d \psi_d < \psi_d$$

This is the parameter that we can reduce with a spatially balanced design

$$g_{2d}\left(\sigma_v^2\right)=\left(1-\gamma_d\right)^2 \mathbf{x}_d^t\left[\sum_d \mathbf{x}_d \mathbf{x}_d^t /\left(\psi_d+\sigma_v^2 b_d^2\right)\right]^{-1} \mathbf{x}_d$$

$$g_{3i}\left(\sigma_v^2\right)=\psi_d^2 b_d^4 /\left(\psi_i+\sigma_v^2 b_d^2\right)^{-3} \bar{V}\left(\hat{\sigma}_v^2\right)$$

A Case Study: Sample Design for Small Area Estimation

- An analytic estimator of the MSE of the FH estimator exist but is not discussed here
- The leading term of MSE is given by $\gamma_i\psi_i$ which shows that the EBLUP estimate can lead to large gains in efficiency over the direct estimate when γ_i is small (or the model variance, σ_v^2 , is small relative to the sampling variance)
- The success of small area estimation, therefore, largely depends on getting good auxiliary data that can lead to a small model variance relative to sampling variance

A Case Study: Sample Design for Small Area Estimation

Theorem 1 (Grafstrom 2012): Let the population consist of two separated regions A and B, such that the within region distances are always less than the distance between units in different regions. If $\sum_{i \in A} \pi_i = n_A$ and $\sum_{i \in B} \pi_i = n_B$, where n_A and n_B are positive integers, then the maximal weight strategy produces samples of fixed sizes, n_A and n_B respectively.

Theorem : If the population is partitioned in $d=1, \dots, D$ domains such that every unit i in d can be included only in the Voronoi polygon of a selected unit j belonging to d , then if $\sum_{i \in d} \pi_i = n_d$ the sample size in each domain will be equal to the size of the domain divided by the average size of its Voronoi polygons:

$n_d^* = n_d / M_d(v_i)$ thus $E(n_d^*) = n_d$ and $V(n_d^*) \approx K(n_d) V_d(v_i)$ (Taylor series). If the sample is spatially balanced then $V_d(v_i) = 0$ and any not planned domain **regularly shaped** will have a controlled sample size.

A Case Study: Sample Design for Small Area Estimation

Objective of the simulation: testing the efficiency of the EBLUP under different SBS, in order to understand the impact of them on the final estimates.

Target area: Italian region “Piemonte”.

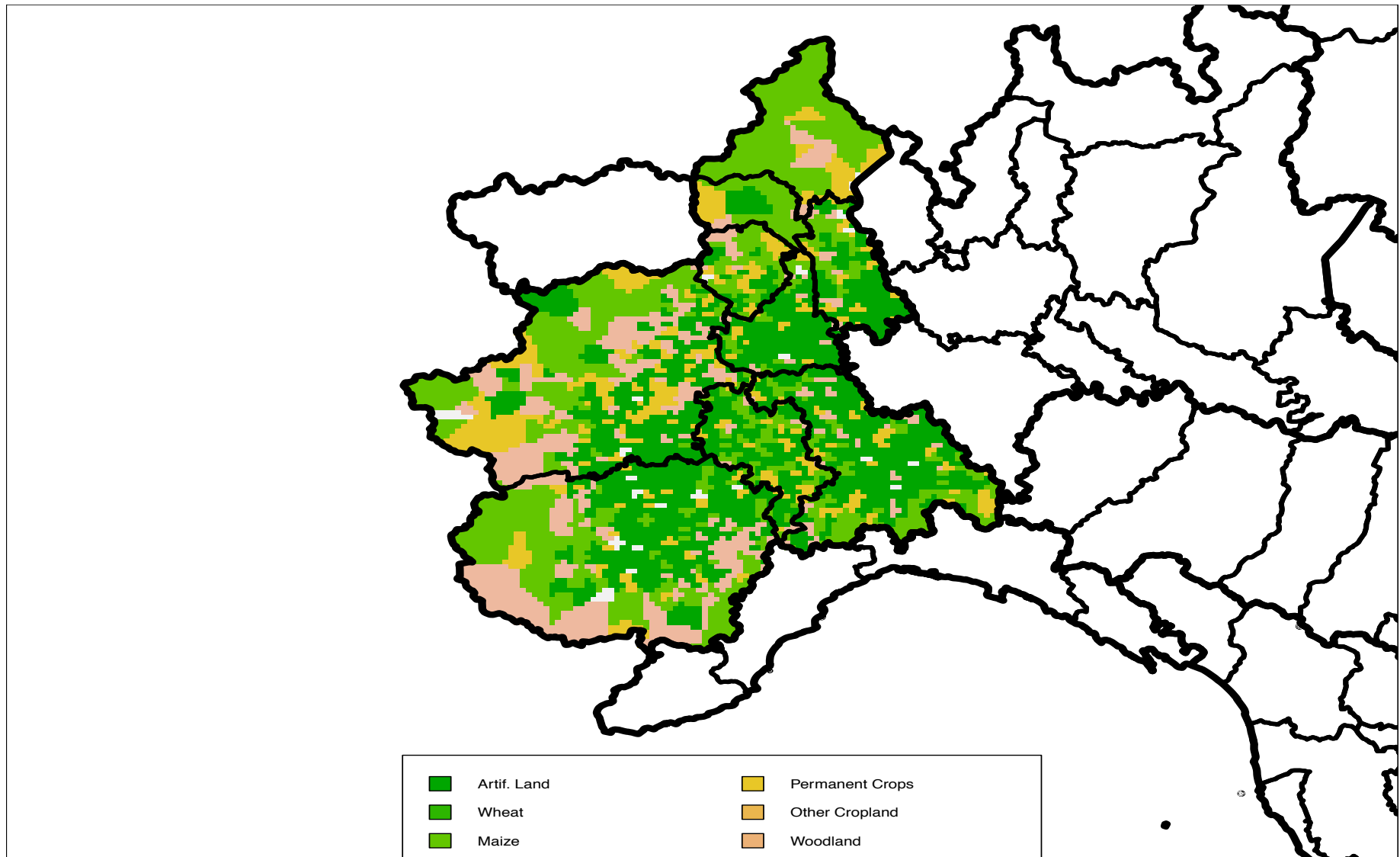
Small areas: 8 provinces, Verbania, Vercelli, Novara, Biella, Torino, Alessandria, Asti, Cuneo.

Target of inference: total of area dedicated to maize at the province level.

Sample data: Land Use/Cover Area Frame Survey (LUCAS).

Auxiliary information: satellite data.

A Case Study: Sample Design for Small Area Estimation



A Case Study: Sample Design for Small Area Estimation

Per each design considered, we selected 1000 samples and computed the relative EBLUPs for the small areas. Next, we computed the RMSE and compared with the RMSE of the EBLUPs computed by SRS.

Designs considered:

- GRTS;
- LPM;
- SCPS;
- PWD with parameter of spread $\beta=10$.

A Case Study: Sample Design for Small Area Estimation

Relative Variability of sample size (with respect to provinces and missing values)										
	n	Torino	Vercelli	Novara	Cuneo	Asti	Alessandria	Biella	Valle d'Aosta	% Missing
Area		6860	2080	1340	6872	1496	3556	892	2276	Values
SRS	100	0,165	0,332	0,423	0,163	0,404	0,248	0,518	0,317	3,540
SRS	300	0,092	0,189	0,239	0,092	0,226	0,138	0,298	0,181	0,000
SRS	600	0,064	0,131	0,165	0,064	0,158	0,096	0,204	0,124	0,000
GRTS	100	0,060	0,219	0,212	0,058	0,250	0,104	0,320	0,119	0,140
GRTS	300	0,027	0,098	0,093	0,024	0,118	0,046	0,156	0,056	0,000
GRTS	600	0,017	0,063	0,059	0,015	0,075	0,028	0,100	0,034	0,000
LPM	100	0,051	0,191	0,178	0,044	0,217	0,084	0,290	0,102	0,050
LPM	300	0,023	0,087	0,080	0,020	0,102	0,037	0,136	0,045	0,000
LPM	600	0,014	0,053	0,049	0,012	0,062	0,023	0,084	0,027	0,000
SCPS	100	0,047	0,183	0,171	0,043	0,200	0,082	0,274	0,092	0,030
SCPS	300	0,021	0,083	0,078	0,018	0,094	0,036	0,126	0,040	0,000
SCPS	600	0,013	0,051	0,047	0,010	0,058	0,021	0,078	0,025	0,000
PWD10	100	0,042	0,167	0,143	0,035	0,190	0,069	0,250	0,079	0,000
PWD10	300	0,019	0,075	0,067	0,016	0,088	0,031	0,118	0,038	0,000
PWD10	600	0,012	0,047	0,042	0,010	0,055	0,019	0,074	0,023	0,000

A Case Study: Sample Design for Small Area Estimation

RMSE HT

	Verbania	Vercelli	Novara	Biella	Torino	Alessandria	Asti	Cuneo
GRTS100	NaN	0.9379392	0.9045098	1.0384114	0.8776712	0.8954728	0.8715355	1.6303230
GRTS300	NaN	0.8759636	0.9303130	0.9520209	0.8928331	1.0050607	0.9278795	0.7700242
GRTS600	NaN	0.8467080	0.8566005	0.9642335	0.7948931	0.9231656	0.8745340	0.7624706
LPM100	NaN	1.0302313	1.0465175	1.1059183	1.0144736	0.9820433	1.0570642	0.9760504
LPM300	NaN	0.9751075	1.0392123	0.9782106	1.0515498	1.1299132	1.0488890	0.8732756
LPM600	NaN	1.0250021	1.0003363	1.0288968	1.1043074	1.0564553	0.9886532	0.8214838
SCPS100	NaN	1.0380086	1.0118318	1.1432814	1.0073035	1.0126819	1.0137090	0.9133247
SCPS300	NaN	1.0093990	1.0063309	1.0097694	1.1105320	1.0888267	1.0611448	0.8446115
SCPS600	NaN	1.0171710	0.9886991	0.9875088	1.0862899	1.0125246	0.9997018	0.8423870
PWD100	NaN	0.8638699	0.8823120	1.0247178	0.8500482	0.8711684	0.9248583	0.8822970
PWD300	NaN	0.7875451	0.8384752	0.8446205	0.8492508	0.9767475	0.8746997	0.7735358
PWD600	NaN	0.7805701	0.8372350	0.8682516	0.7461031	0.9040853	0.8455749	0.7400753

A Case Study: Sample Design for Small Area Estimation (EBLUP RRMSE)

	Verbania	Vercelli	Novara	Biella	Torino	Alessandria	Asti	Cuneo
GRTS100	0.6360868	1.5635796	0.9888364	0.6794555	1.4223646	0.8675475	1.5691916	1.5296811
GRTS300	0.8353760	0.9034418	1.0170633	0.9447852	0.6173188	1.2353202	0.8641661	0.5553207
GRTS600	0.9346582	0.6455000	0.8420095	0.9188386	0.4561037	1.1307870	0.6653687	0.4170726
LPM100	0.8859967	1.2638340	1.1547011	0.9282241	1.0655236	1.2626772	1.2646547	0.9607005
LPM300	0.9325531	0.8393133	1.3136664	1.0491140	0.6977828	1.5577502	0.8550787	0.5814019
LPM600	1.0621174	0.5987429	1.0063838	1.0596980	0.5833469	1.3774114	0.5750544	0.4045299
SCPS100	0.8796000	1.2495079	1.0724639	0.9267192	1.0214755	1.2282016	1.1788845	0.9027995
SCPS300	0.9661380	0.8831110	1.2819012	1.1082960	0.6821520	1.5427569	0.8075880	0.5480727
SCPS600	1.1021186	0.6043376	0.9777239	1.0584320	0.5777138	1.3890846	0.5685249	0.4033101
PWD100	0.7943272	1.0715450	0.8741273	0.8035085	0.9479421	1.0528199	1.2049230	0.8970312
PWD300	0.7985164	0.8704397	0.9690116	0.8564637	0.6110313	1.2165179	0.8545213	0.5523522
PWD600	0.9059844	0.6511020	0.7414879	0.8171829	0.4222790	1.0944463	0.6687905	0.4105362

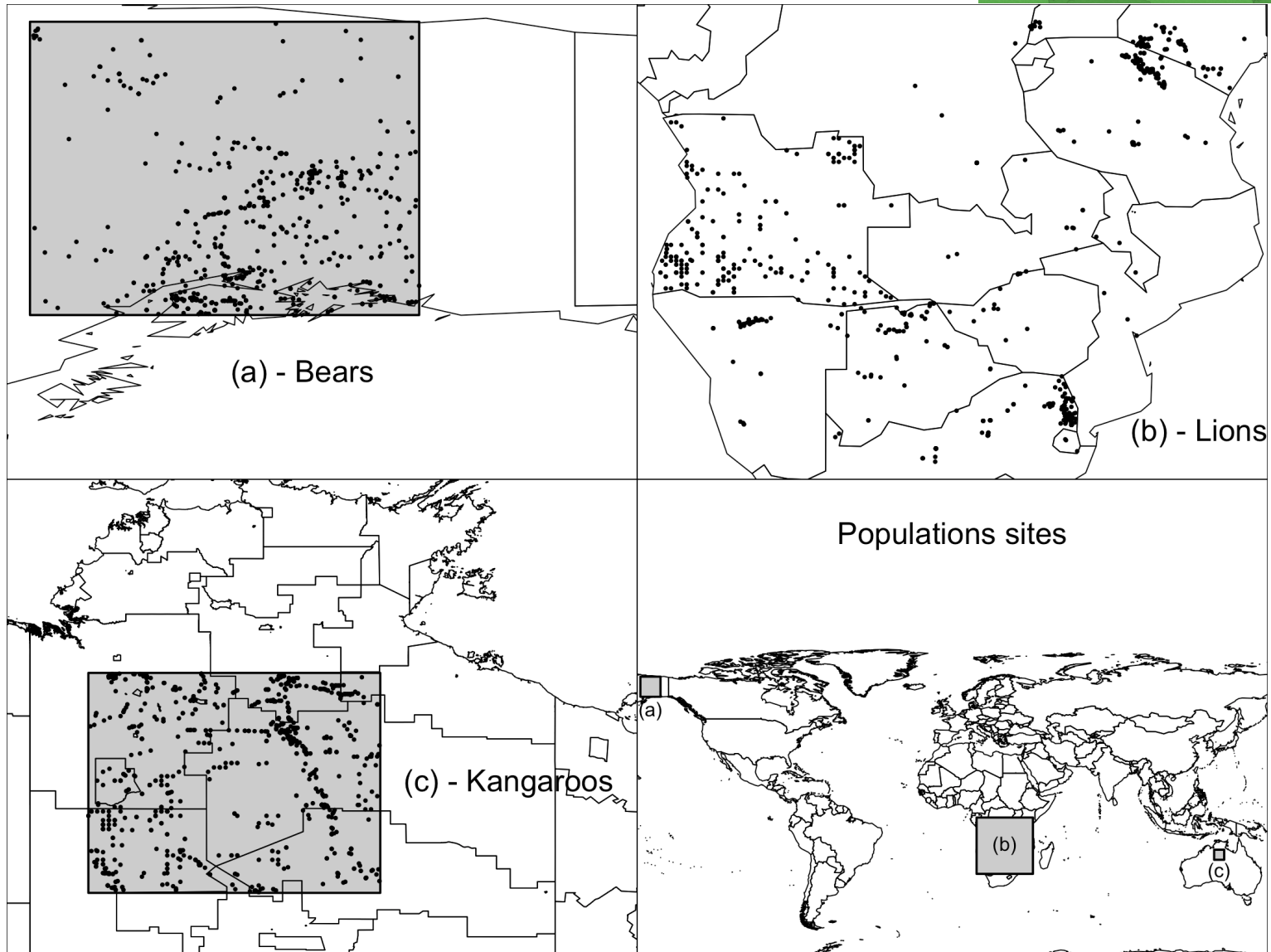
A Case Study: Sample Design for Small Area Estimation

- SBS: taking into account the spatial information when selecting a sample.
- SAE: face the challenge of obtain reliable estimates for small areas (likely with few observations) through the use of a composite estimator.
- SBS + SAE: the results show a possible gain in the efficiency of the final estimates if SBS and SAE are used together. This is due to the possible reduction of the first term of the MSE of the EBLUP.

A Case Study: Wild Animals



Global Biodiversity
Information Facility
Free and Open Access to Biodiversity Data



A Case Study: Wild Animals

Systematic and Maximal Stratification (also called TSS or 1 unit per stratum) are very difficult to be extended to regions not box shaped. At least it is very difficult to control the sample size n and to stratify the region in n non overlapping strata. The aim is to find a set of alternative *spatially balanced designs* that, even if they require a discretization of the region, are more flexible and guaranty approximately the same efficiency. This experiment is thus conducted in the most favourable situation for SYS and TSS: a square region. If another design performs as well as SYS and TSS in this situation, and it is more flexible to other regions with irregular shape, it can be preferred.

A Case Study: Wild Animals

3 Populations

a – Bears (N=767)

b – Lions (N=873)

c – Kangaroos (N=1069)

10.000 samples

3 estimates

Perfect Detection

Known Detection Function

Unknown Detection Function

3 Sample Size

Point Transect (n = 49, 100, 144)

$N = 50 \times 50 = 2500$

Line Transect (n = 5, 10, 15)

$N = 100$

10 Sample Selection Algorithms

srs, sys, st1, hal, cub1, cub2, scps, lpm, lcb1, lcb2

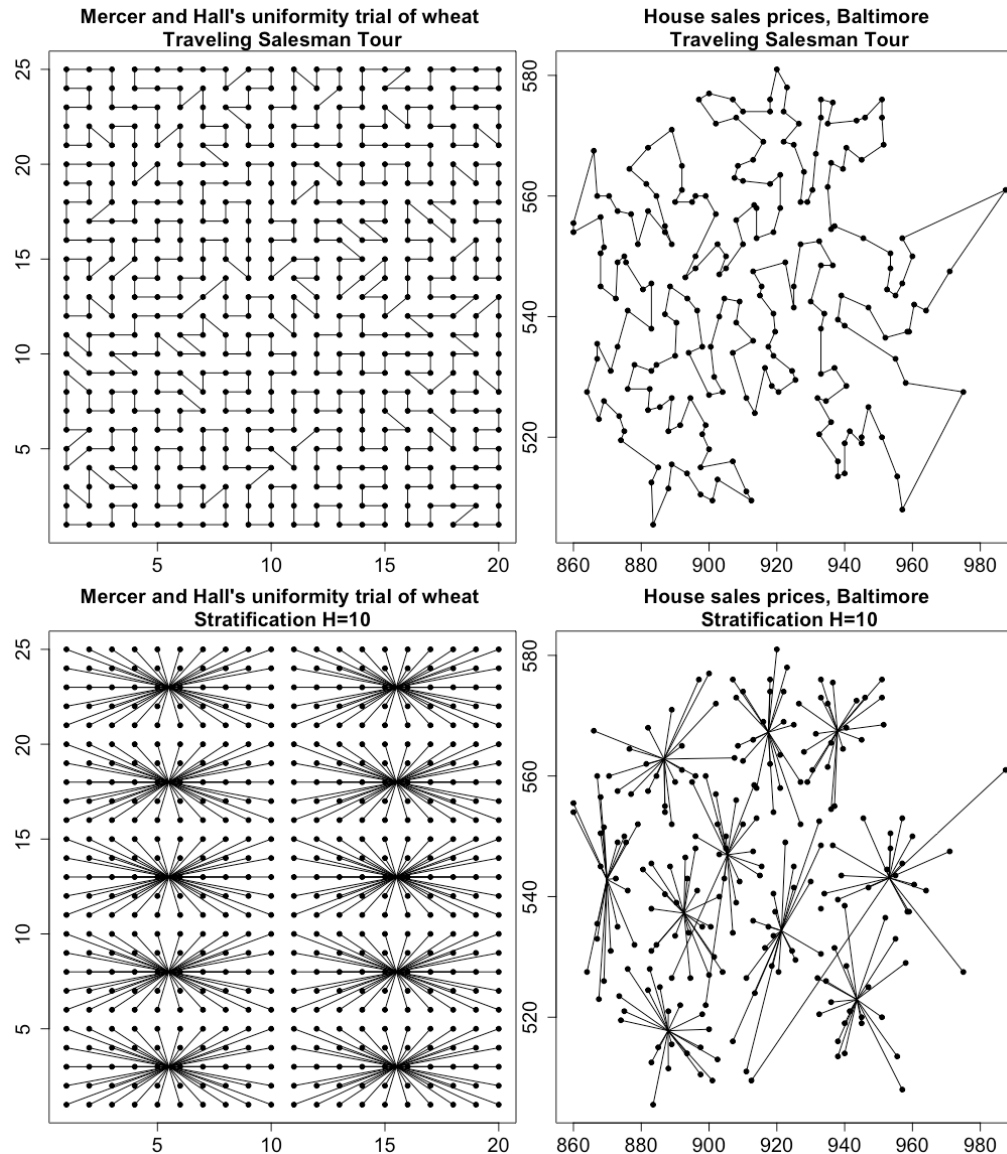
The detection was simulated according to an Half-normal model of parameter $\alpha = 0.01$

The following tables reports the relative RMSE of abundance estimates using the SRS as benchmark

A Case Study: Wild Animals

	Kangaroos			Lions			Bears		
Method	n=49	n=100	n=144	n=49	n=100	n=144	n=49	n=100	n=144
sys	0,93	0,64	0,53	0,88	0,64	0,52	0,94	0,58	0,49
st1	0,93	0,57	0,44	0,94	0,47	0,35	0,92	0,60	0,42
BAL	0,90	0,60	0,48	0,92	0,56	0,46	0,95	0,56	0,44
cub1	0,99	0,56	0,45	0,99	0,53	0,43	0,98	0,55	0,43
cub2	1,00	0,58	0,53	0,99	0,53	0,52	0,97	0,54	0,47
scps	0,94	0,58	0,53	0,92	0,53	0,52	0,96	0,55	0,47
lpm	0,96	0,58	0,47	0,90	0,53	0,41	0,96	0,54	0,43
lcb1	0,94	0,58	0,47	0,91	0,52	0,42	0,94	0,55	0,44
lcb2	0,91	0,66	0,47	0,90	0,63	0,41	0,92	0,59	0,45

A Case Study: Regular Grid vs Irregular Points



A Case Study: Regular Grid vs Irregular Points

	n	Mercer-Hall		Baltimore		n	Mercer-Hall		Baltimore	
		RMSE (Relative to SRS)	SBI	RMSE (Relative to SRS)	SBI		RMSE (Relative to SRS)	SBI	RMSE (Relative to SRS)	SBI
GRTS	10	0.927	0.142	0.860	0.173	50	0.875	0.111	0.763	0.179
CUBE	10	0.963	0.199	0.886	0.221	50	0.955	0.278	0.862	0.313
LPM	10	0.919	0.101	0.823	0.126	50	0.862	0.085	0.706	0.148
SCPS	10	0.923	0.086	0.817	0.112	50	0.846	0.074	0.656	0.151
DBSS	10	0.919	0.088	0.779	0.114	50	0.847	0.073	0.679	0.145
KM	10	0.910	0.093	0.833	0.119	50	0.885	0.096	0.696	0.186
TSP	10	0.942	0.111	0.852	0.117	50	0.878	0.103	0.672	0.155
KMSB	10	0.906	0.092	0.829	0.118	50	0.874	0.109	0.751	0.197
TSPSB	10	0.921	0.108	0.832	0.118	50	0.879	0.112	0.712	0.170
KMDBSS	10	0.914	0.082	0.807	0.108	50	0.876	0.091	0.674	0.179
TSPDBSS	10	0.926	0.094	0.793	0.107	50	0.888	0.097	0.652	0.151

Computational Burden

N	n	GRTS	CUBE	SCPS	LPM1	LPM2	PWD	HPWD	SCPS_C	LPM1_C	LPM2_C
1000	100	0,182	0,152	3,114	8,348	2,008	0,243	0,014	0,058	0,013	0,007
1000	300	1,941	0,158	3,094	8,457	2,010	0,374	0,039	0,058	0,013	0,007
1000	600	57,008	0,145	2,962	8,137	1,960	0,594	0,075	0,058	0,013	0,007
1500	100	0,116	0,220	6,728	17,698	4,393	0,376	0,019	0,135	0,028	0,015
1500	300	0,848	0,220	6,665	16,662	4,161	0,598	0,059	0,133	0,028	0,015
1500	600	23,195	0,220	6,668	17,627	4,375	0,952	0,112	0,134	0,028	0,015
2000	100	0,104	0,293	11,920	31,587	7,757	0,504	0,026	0,246	0,050	0,026
2000	300	1,088	0,293	11,848	31,586	7,682	0,791	0,074	0,246	0,050	0,026
2000	600	5,267	0,286	11,718	31,084	7,661	1,264	0,147	0,245	0,050	0,026
2500	100	0,115	0,361	18,511	49,659	12,036	0,642	0,031	0,392	0,078	0,041
2500	300	1,108	0,358	18,304	49,919	12,051	1,019	0,094	0,392	0,078	0,041
2500	600	5,021	0,365	18,318	49,666	12,053	1,652	0,185	0,392	0,078	0,041
3000	100	0,134	0,431	26,623	71,000	17,255	0,769	0,038	0,574	0,112	0,059
3000	300	1,418	0,426	26,412	71,197	17,343	1,242	0,115	0,573	0,113	0,059
3000	600	4,688	0,450	26,532	66,318	16,284	1,972	0,223	0,564	0,113	0,059

Thank you for your attention!