

NFI parameter estimation

Selection of practical topics based on Czech NFI2 experience

Radim Adolt

ÚHÚL Brandýs nad Labem, Czech Republic



Forest Inventory Statistics, seminar organized by WSL
Birmensdorf, 25.-26. March 2015



- 1 Accuracy of one- and two-phase estimators in a typical NFI design
 - Problem description
 - Simulation study
 - Proposed solutions
- 2 Confidence intervals and central limit theorem
 - Problem description
 - Simulation study
 - Proposed solutions
- 3 Appendix



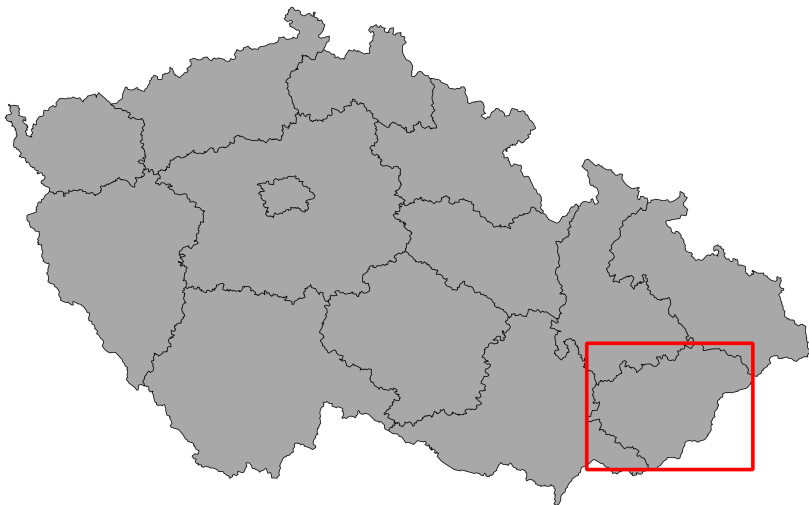
Problem description

- **(quasi-)systematic** (spatially restricted) designs used by most NFIs
- **conservative URS based estimators** regularly applied
- **overestimation of variance** due to spatial correlation
- two-phase variance estimation based predominantly on **residuals** having **reduced spatial correlation**
- **different levels of conservativeness** between one- and two-phase
- **true effect of two-phase approach usually overestimated**

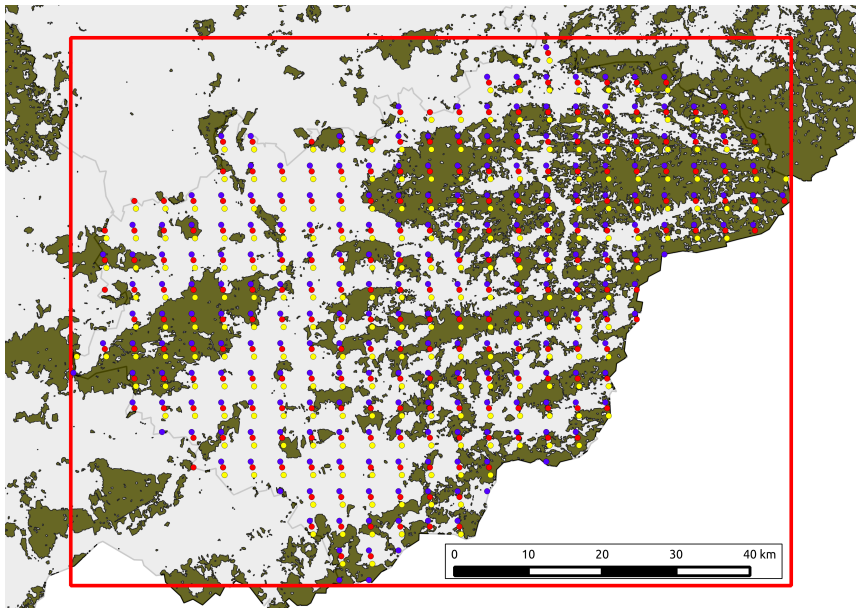


Area for the simulation study

Study area within Czech Republic - Zlinsky kraj (NUTS3 level)

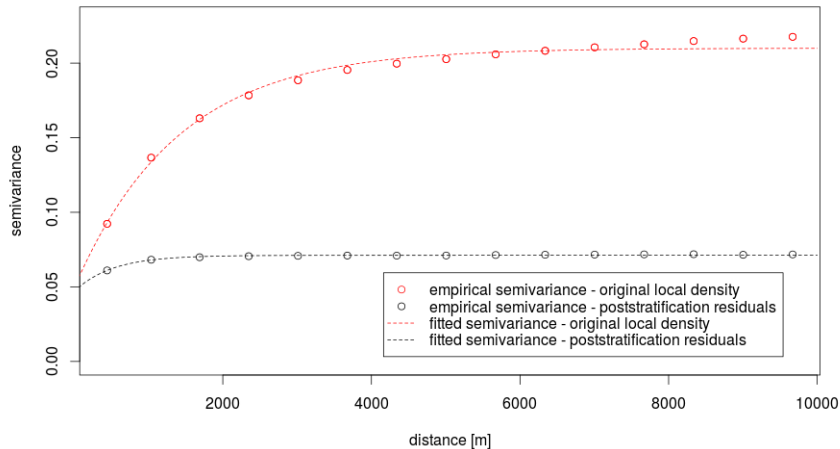


Demonstration of three CSS samples



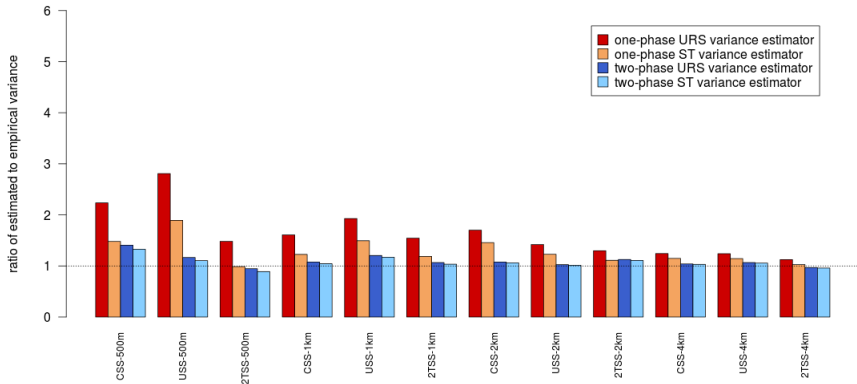
Variography - based on simulated URS samples

Variograms for one and two phase total forest area estimation,
JRC Forest Type Map (2006) used as real population



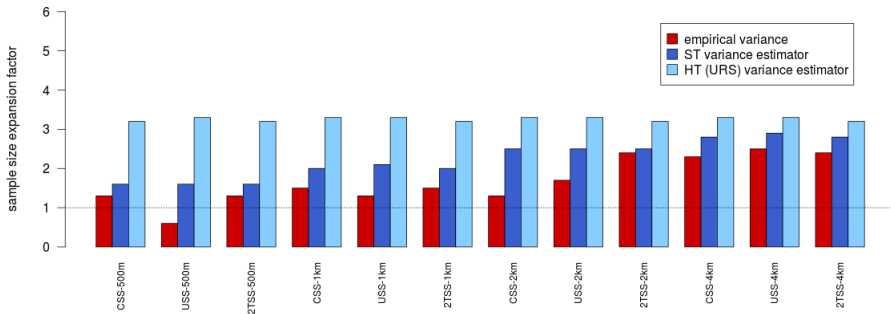
Estimated and empirical variances (one- and two-phase)

Ratio of URS and ST variance estimates to empirical variance, total forest area, one- and two-phase estimators, various designs and resolutions (NUTS3/cz072).



Estimated and empirical effects of two-phase approach

Empirical, URS and ST sample size expansion factors, total forest area, various designs and resolutions (NUTS3/cz072).



Proposed solutions

- **Do not trust URS variance estimators blindly!**
- **Use alternative variance estimators** - possibly suited for particular design [Cordy, 1993, Cordy and Thompson, 1995, Stevens and Olsen, 2003, Cooper, 2006]...
- **Explore spatial correlation** for very important target parameters (limited number of such parameters).
- The problem seems to be reduced in case of **ratio estimation** - based on Taylor approximation a residual variable is constructed and used for variance estimation in one-phase setting.
- Lower risk is expected for **means over homogeneous domains** with relatively small range of spatial correlation - e.g. per hectare growing stock calculated as the arithmetic mean of local density over inventory points belonging to stocked area.

Problem description

- probability distribution of an estimator is **assumed to be normal**
- most **local densities (untransformed) are not normally distributed** - very often they use to be right skewed
- estimators are **sums of local densities**, additionally multiplied by other quantities (often constants)
- normality assumption is justified by **Central Limit Theorem (CLT)**
- URS estimators are sums of **equally distributed and independent random variables**



Problem description

- for non-normally distributed estimator **coverage of the normal based confidence interval does not fit its nominal value**
- standard CLTs hold for **“reasonable” sample sizes** and URS design
 - **CSS** (Centric Systematic Sampling)- **lack of independence**, strong and weak mixing theorems applicable see Iachan [1983]
 - **TSS** (Tesselated Stratified Sampling) - local density in each of the fixed strata has potentially **different distribution** (e.g. mean and/or variance), convergence to normality reported in Barabesi [2003]
- **no general rule for minimum sample size** to safely trust in normality convergence due to CLT



Empirical coverage, Prague county (124 tracts, 248 plots)

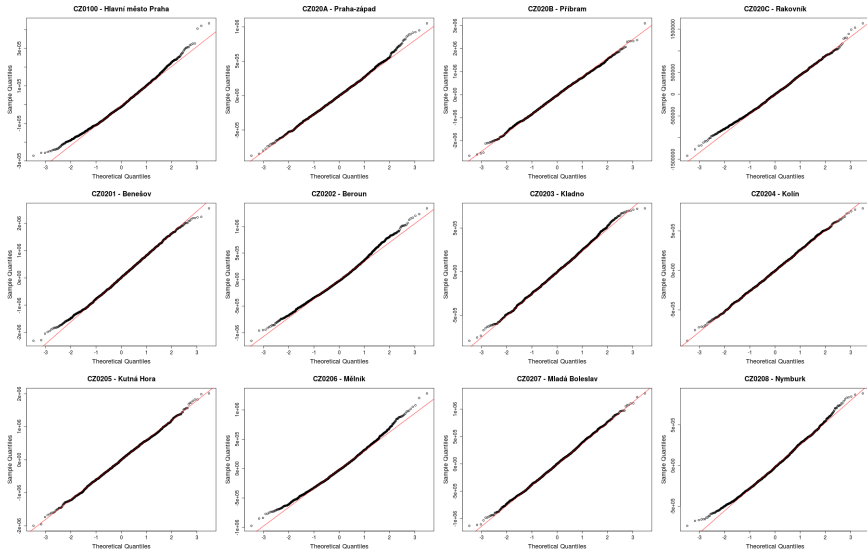
Table C.9: One-phase total growing stock [m³] estimates, NUTS4 level (nruns=2000)

geographic domain	parameter	emean	bias	pbias [%]	rbias [%]	estd	rsht	rsst	echt	ecst
CZ0100 - Hlavní město Praha	all_stock	838765	756	0.090	0.450	168040.168	1.294	1.202	0.975	0.963
CZ0100 - Hlavní město Praha	brd_stock	491539	3267	0.669	2.542	128534.062	1.095	1.061	0.953	0.949
CZ0100 - Hlavní město Praha	con_stock	347224	-2511	-0.718	-2.414	104005.123	1.178	1.053	0.939	0.920
CZ020A - Praha-západ	all_stock	3097097	4949	0.149	1.499	309938.991	1.420	1.302	0.994	0.989
CZ020A - Praha-západ	brd_stock	823348	3604	0.440	2.855	126232.292	1.212	1.146	0.974	0.966
CZ020A - Praha-západ	con_stock	2234345	942	0.042	0.341	276367.190	1.357	1.269	0.985	0.978
CZ020B - Příbram	all_stock	17521796	-32344	-0.184	-3.928	823385.786	1.322	1.122	0.989	0.967
CZ020B - Příbram	brd_stock	1369929	1885	0.138	0.921	20417.708	1.036	0.974	0.956	0.944
CZ020B - Příbram	con_stock	16151848	-34229	-0.211	-4.335	789675.302	1.316	1.121	0.987	0.970
CZ020C - Rakovník	all_stock	6496045	6253	0.096	1.251	499961.456	1.227	1.097	0.987	0.973
CZ020C - Rakovník	brd_stock	1672512	2747	0.165	1.278	214967.559	1.208	1.068	0.971	0.950
CZ020C - Rakovník	con_stock	4823531	3506	0.073	0.821	427283.635	1.202	1.123	0.985	0.973
CZ0201 - Benešov	all_stock	10944568	22555	0.207	2.721	828956.827	1.134	1.116	0.976	0.974
CZ0201 - Benešov	brd_stock	1099229	1894	0.173	1.122	168761.827	1.069	1.031	0.954	0.946
CZ0201 - Benešov	con_stock	9845325	20661	0.210	2.630	785645.291	1.140	1.124	0.979	0.972
CZ0202 - Beroun	all_stock	4563123	5120	0.112	1.212	422565.930	1.237	1.100	0.982	0.970
CZ0202 - Beroun	brd_stock	1449544	-251	-0.017	-0.140	178989.106	1.261	1.168	0.979	0.971
CZ0202 - Beroun	con_stock	3113563	5370	0.173	1.473	364707.328	1.185	1.035	0.981	0.953
CZ0203 - Kladno	all_stock	2858652	-6755	-0.236	-2.239	301738.252	1.496	1.185	0.992	0.973
CZ0203 - Kladno	brd_stock	869241	-6053	-0.692	-4.280	141436.224	1.397	1.138	0.986	0.963
CZ0203 - Kladno	con_stock	1989409	-701	-0.035	-0.280	250464.033	1.446	1.229	0.988	0.976
CZ0204 - Kolín	all_stock	2030474	-313	-0.015	-0.105	297312.688	1.333	1.190	0.978	0.969
CZ0204 - Kolín	brd_stock	629879	2601	0.415	1.826	142435.828	1.231	1.193	0.965	0.964
CZ0204 - Kolín	con_stock	1400592	-2914	-0.208	-1.131	257679.563	1.249	1.116	0.968	0.955
CZ0205 - Kutná Hora	all_stock	6121317	-13882	-0.226	-2.169	639876.719	1.207	1.104	0.975	0.956
CZ0205 - Kutná Hora	brd_stock	599664	-4675	-0.774	-3.439	135949.065	1.078	1.073	0.928	0.923
CZ0205 - Kutná Hora	con_stock	5521648	-9206	-0.166	-1.532	600903.401	1.240	1.124	0.980	0.967
CZ0206 - Mělník	all_stock	2697408	-12143	-0.448	-2.714	447337.964	0.946	0.801	0.940	0.884
CZ0206 - Mělník	brd_stock	884647	-5277	-0.593	-2.560	206170.287	0.905	0.899	0.911	0.906
CZ0206 - Mělník	con_stock	1812746	-6866	-0.377	-2.133	321814.022	1.072	0.880	0.954	0.909
CZ0207 - Mladá Boleslav	all_stock	5760535	6627	0.115	1.472	450063.898	1.399	1.226	0.994	0.984
CZ0207 - Mladá Boleslav	brd_stock	1515045	7127	0.473	3.023	235748.757	1.080	1.040	0.960	0.955

Continued on next page

Normal Q-Q plots for simulated estimates

Q-Q plots for normal distribution, one-phase total estimator, CON_STOCK, NUTS4 level



Proposed solutions

- **minimum sample size rule** by Cochran [1977, equations 2.69, 2.70, page 42] - tested by simulated sampling followed by one-phase estimation of total growing stock (Czech NFI1 sampling design - quasi-systematic), if minimum sample size was fulfilled, minimum coverage (94%) was almost always fulfilled, otherwise it was met in more than 80% of cases (conservative estimators of variance)
- **ratios and two-phase estimation** - normality more frequently reached because residuals use to be symmetrically distributed



Minimum sample size after Cochran [1977]

Cochran recommended following:

$$n_{min} > 25G_1^2, \quad (1)$$

where G_1 is Fisher's skewness and n_{min} is the minimum sample size. Applicable for data that deviate from normality by skewness only. Equation (1) is designed to reach at least 94% coverage when 95% is nominally required.



- L. Barabesi. A monte carlo integration approach to horvitz-thompson estimation in replicated environmental designs. *International Journal of Statistics*, LXI:355–374, 2003.
- W. G. Cochran. *Sampling Techniques*. Willey series in probability and mathematical statistics - applied. John Willey & Sons, 1977.
- C. Cooper. Sampling and variance estimation on continuous domains. *Environmetrics*, 17:539–553, 2006.
- B. Cordy, C. and M. Thompson, C. An application of the deterministic variogram to design-based variance estimation. *Mathematical Geology*, 27:173–205, 1995.
- C. B. Cordy. An extension of the horwitz-thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*, 18:353–362, 1993.



R. Lachan. Asymptotic theory of systematic sampling. *The Annals of Statistics*, 11(3):959–969, 1983.

D. L. Jr. Stevens and A. R. Olsen. Variance estimation for spatially balanced samples of environmental resources. *Environmetrics*, 14: 593–610, 2003.

