PhD. Dissertation Capstone

Uncertainty Evaluation of Sample-Based Area Estimates in Land Cover Monitoring: Improved Methods for Estimating Confidence Intervals and Total Variance

Dingfan Xing

Major Professor: Dr. Stephen V. Stehman

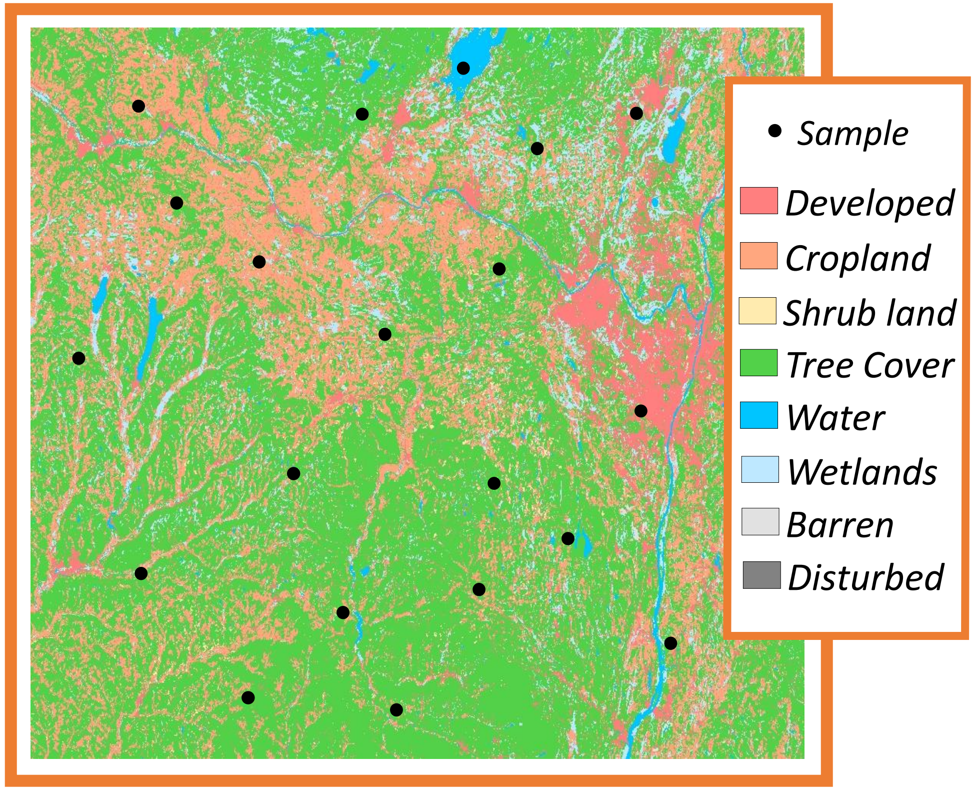
23rd October 2023

ESF Bray 324

State University of New York
COLLEGE OF ENVIRONMENTAL SCIENCE AND FORESTRY
Graduate Program in Environmental Science (GPES)
Sustainable Resources Management (SRM)

Background

Sample Data



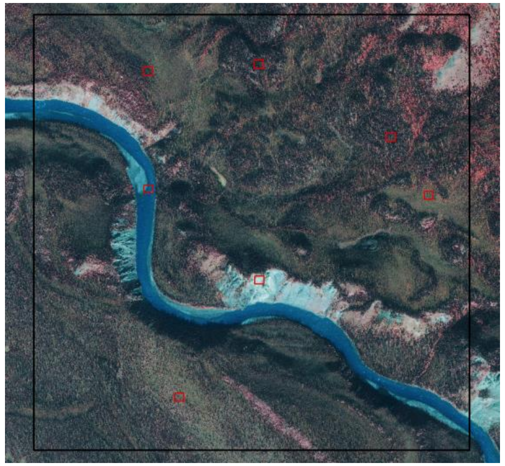
Sample unit :

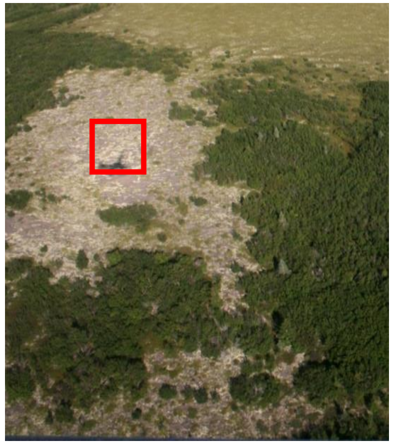
- Pixels

Sampling method:

- Simple random sampling (SRS) & Stratified random sampling (STR)

Reference Data





Area Estimates of Land Cover

Class#	Class	Area(2000)
1	Water	2.3%
2	Developed	26.0%
3	Disturbed	0.0%
4	Barren	0.0%
5	TreeCover	51.7%
6	Grass&Shrub	3.0%
7	Cropland	13.7%
8	Wetland	3.0%

Background

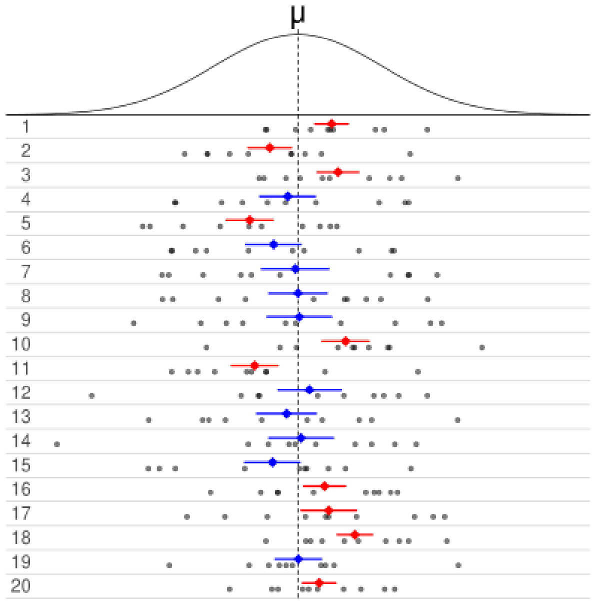
❖ *Confidence Intervals for a Proportion: Rare Classes*

$$CI_W = \hat{p} \pm z * SE(\hat{p})$$

Undercoverage Problem

Alternative Intervals?

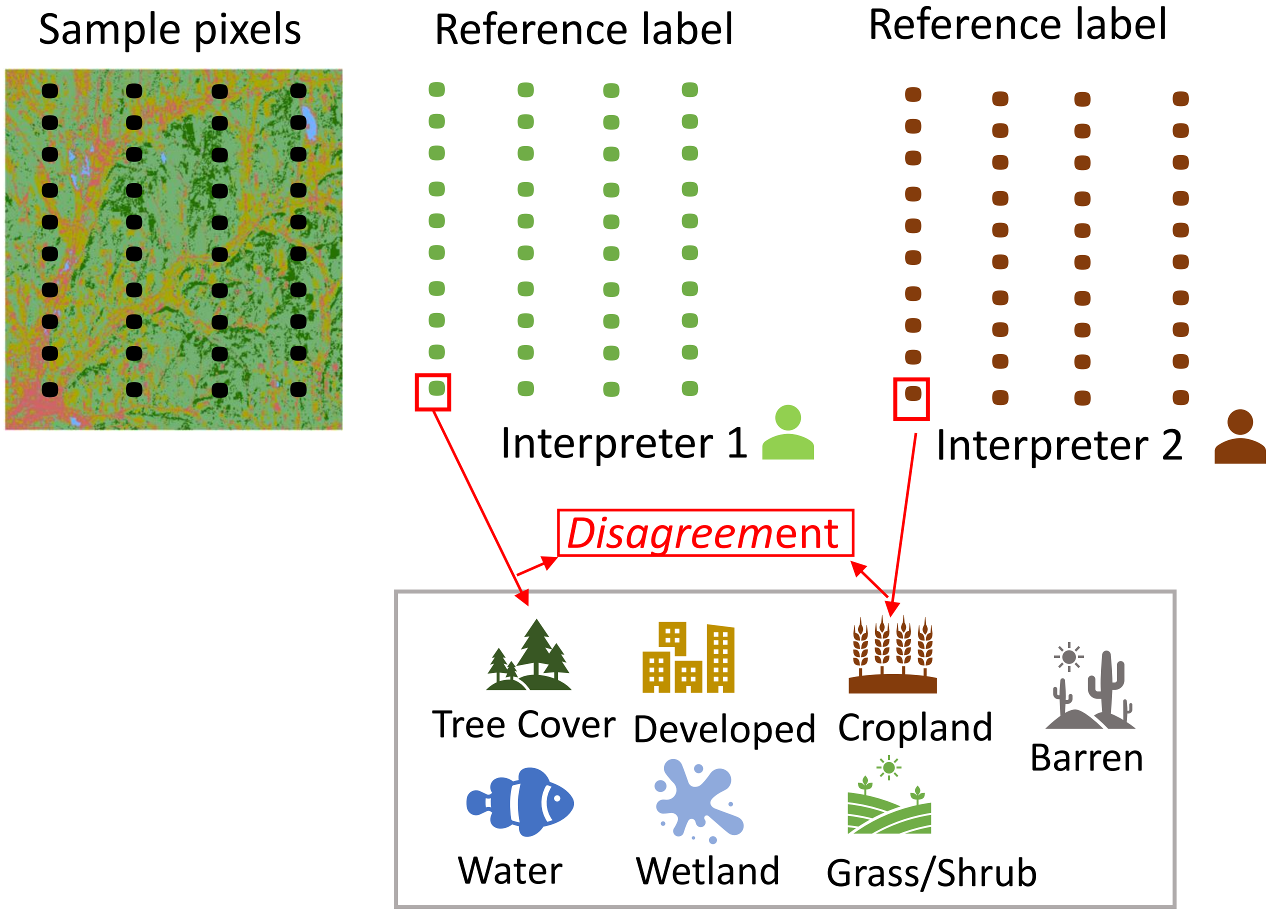
Factors Affecting Coverage When H>2?



Manuscript 1. Factors Affecting Confidence Interval Coverage for Proportion of Area of a Rare Class in Stratified Random Sampling

Background

❖ *Total Variance Estimation*



Sampling Variance

+

Measurement Variance

⇓

Total Variance

Total Variance ?

Incorporating Reference Data Variability to Area Estimates of Land Cover

Manuscript 2 & 3

Background

❖ ***Three Manuscripts***

❖ ***Confidence Intervals for a Proportion: Rare Classes***

Manuscript 1. Factors Affecting Confidence Interval Coverage for Proportion of Area of a Rare Class in Stratified Random Sampling

❖ ***Total Variance Estimation***

Manuscript 2. Using Interpenetrating Subsampling to Incorporate Interpreter Variability into Estimation of the Total Variance of Land Cover Area Estimates

Manuscript 3. Applications of Total Variance Estimation Incorporating Reference Data Variability to Area Estimates of Land Cover

Manuscript 1

Confidence Intervals for a Proportion: Rare Classes

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Introduction

Wald interval

$$CI_W = \hat{p} \pm z * SE(\hat{p})$$

General form

$$= \hat{p} \pm z * \sqrt{\hat{p}(1-\hat{p})/(n-1)}$$

For simple random sampling

Wald interval tends to undercover when proportion of area *p* is small (i.e., less than 5%)

Alternative intervals

✓ Wilson Interval

•
•
•

Modifications for stratified random sampling

✓ Effective sample size (*neff*) (Franco et al. (2019))

neff method is based on estimating effective sample size

✓ *sumstrat* (Stehman and Xing (2022))

sumstrat is based on summing stratum-specific confidence bounds

However…

Franco et al. (2019) focused much of their analysis on cluster sampling

Stehman and Xing (2022) limited the investigation to H=2

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

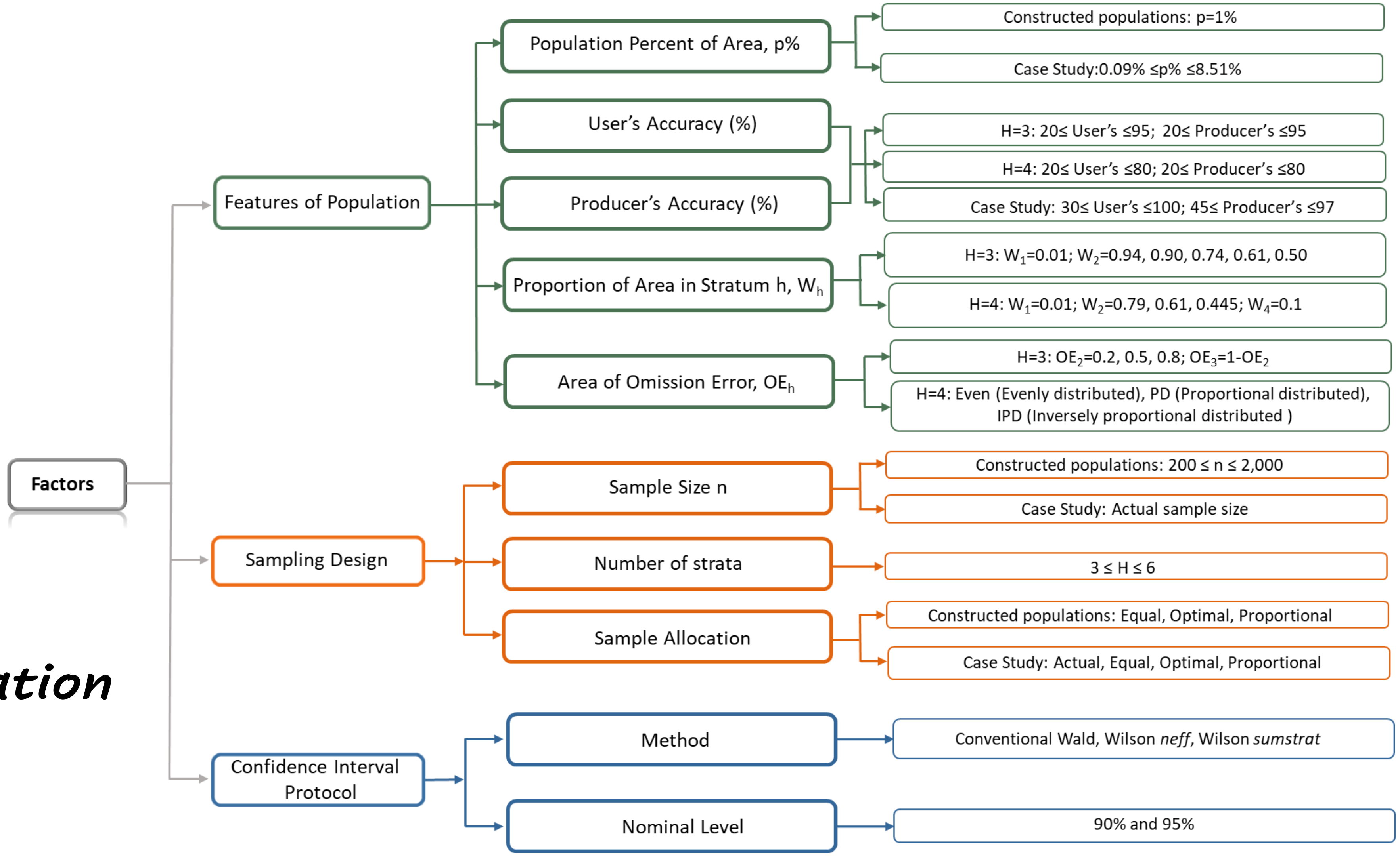
❑ Methods

Populations

- ✓ H=3
- ✓ H=4
- ✓ Case study

Potential Factors

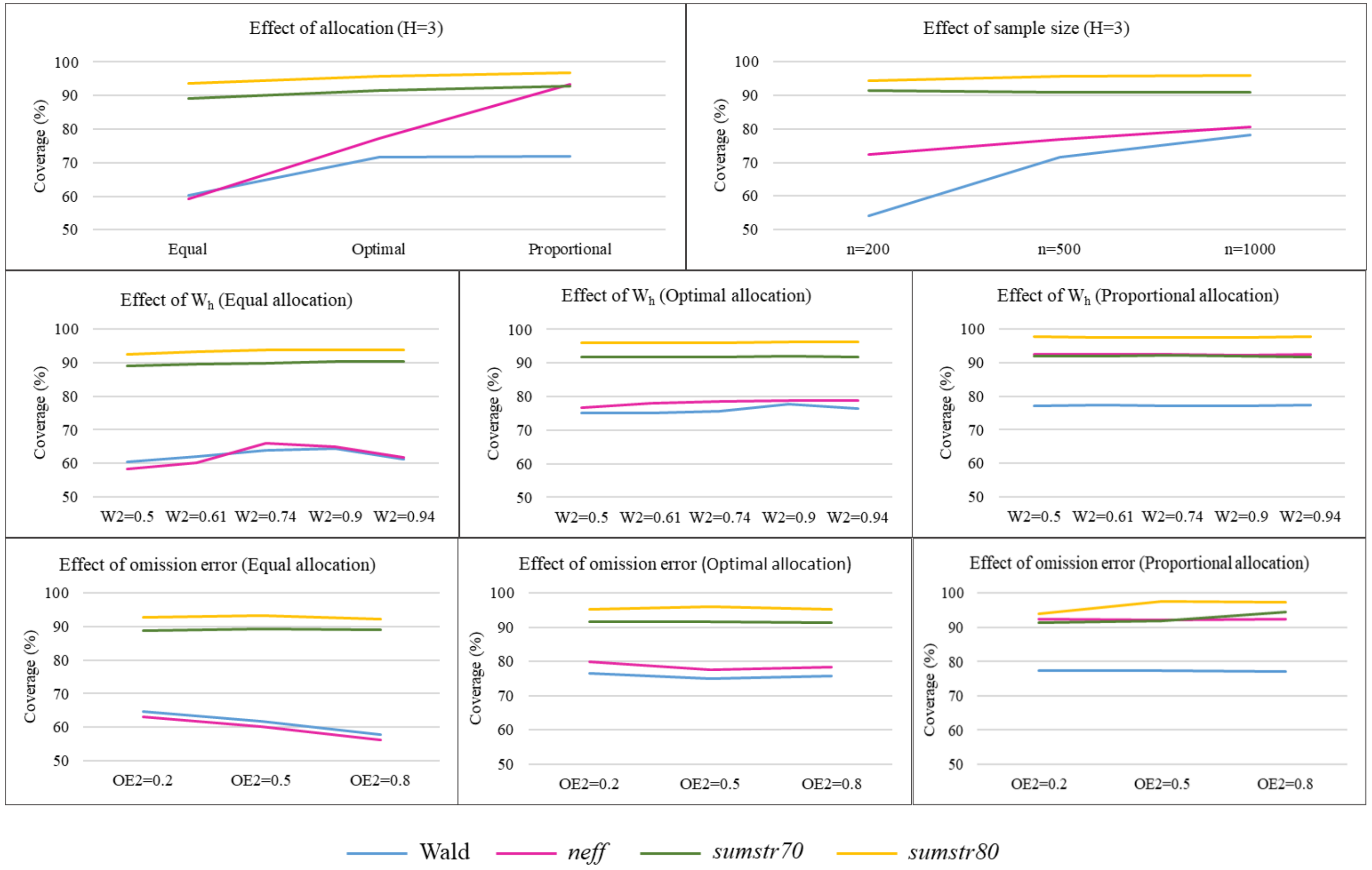
Monte-Carlo Simulation



Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

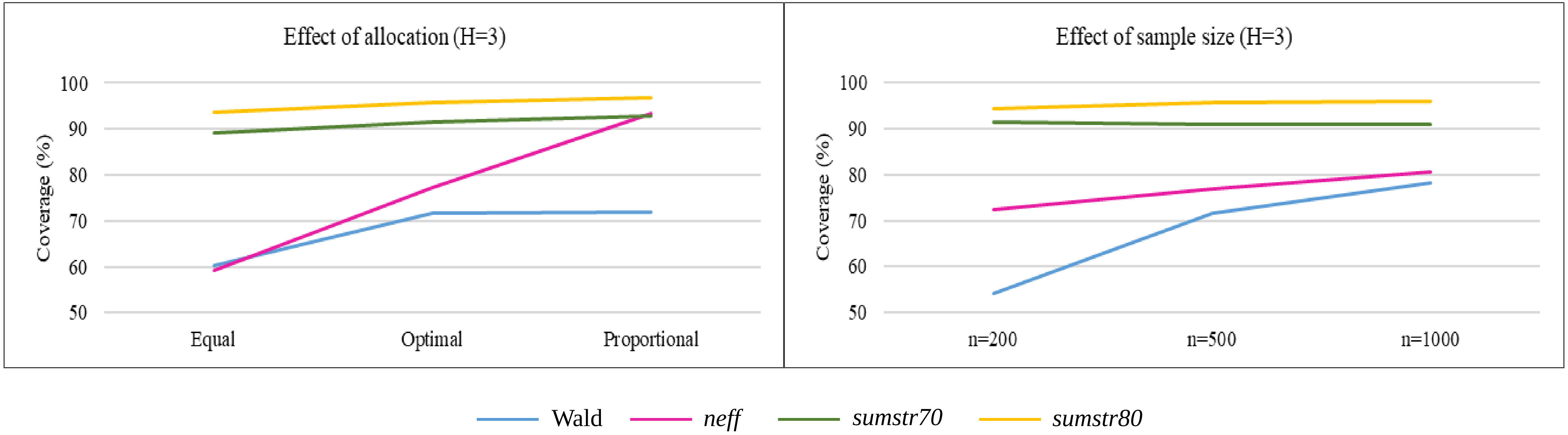
- ✓ Confidence Intervals
- ✓ Allocation
- ✓ Sample size
- ✓ Impact of stratum weight
- ✓ Impact of distribution of omission error



H=3

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

- Results
  - ✓ Allocation
  - ✓ Sample size

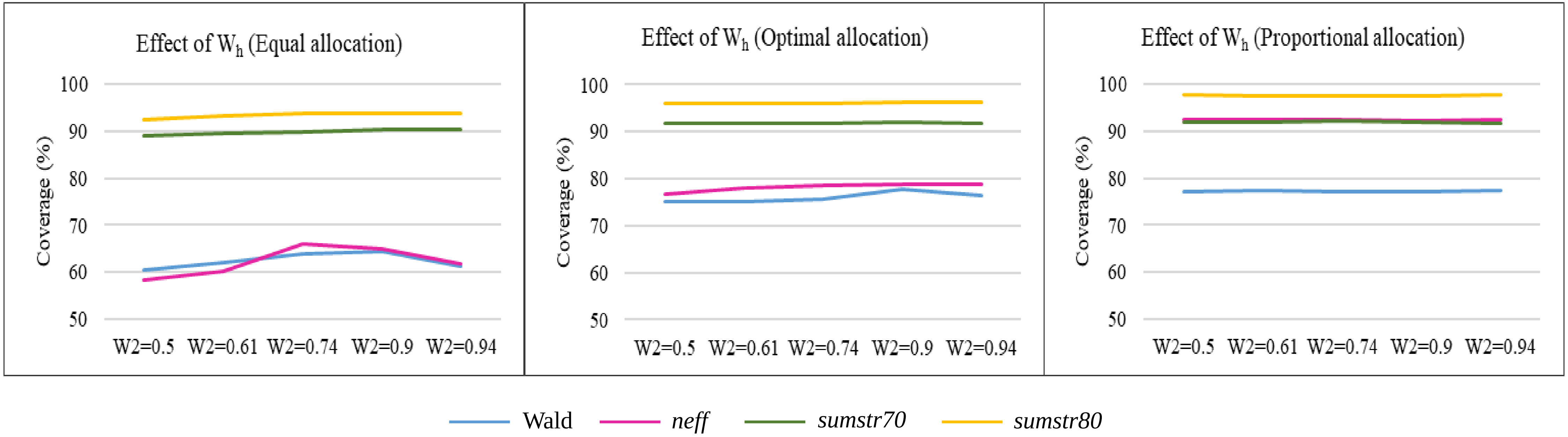


H=3

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

✓ Impact of stratum weight

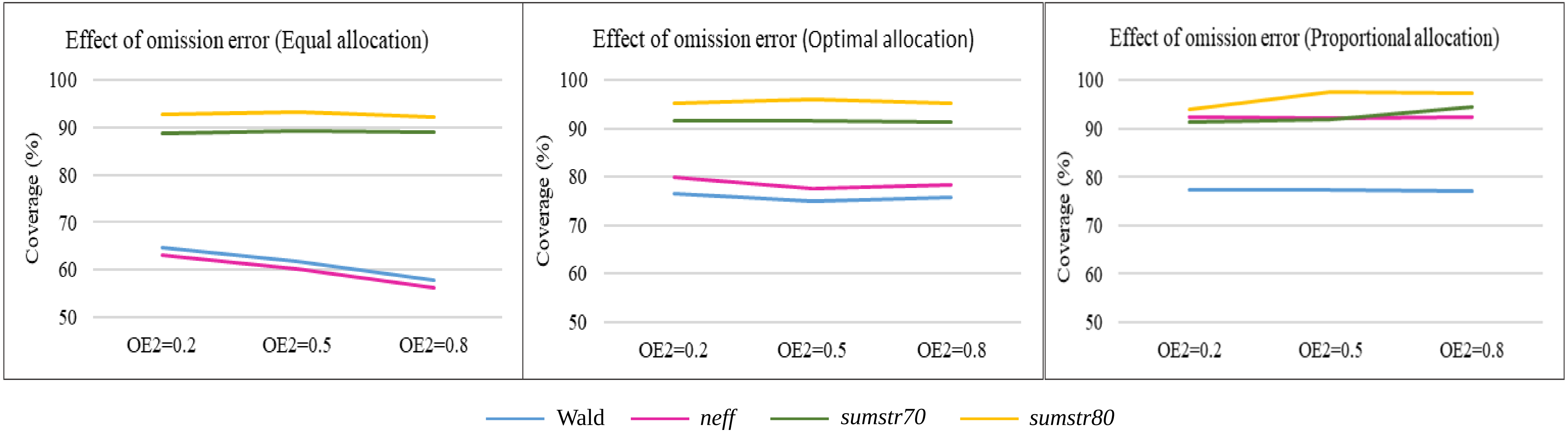


H=3

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

✓ Impact of distribution of area of omission error to strata

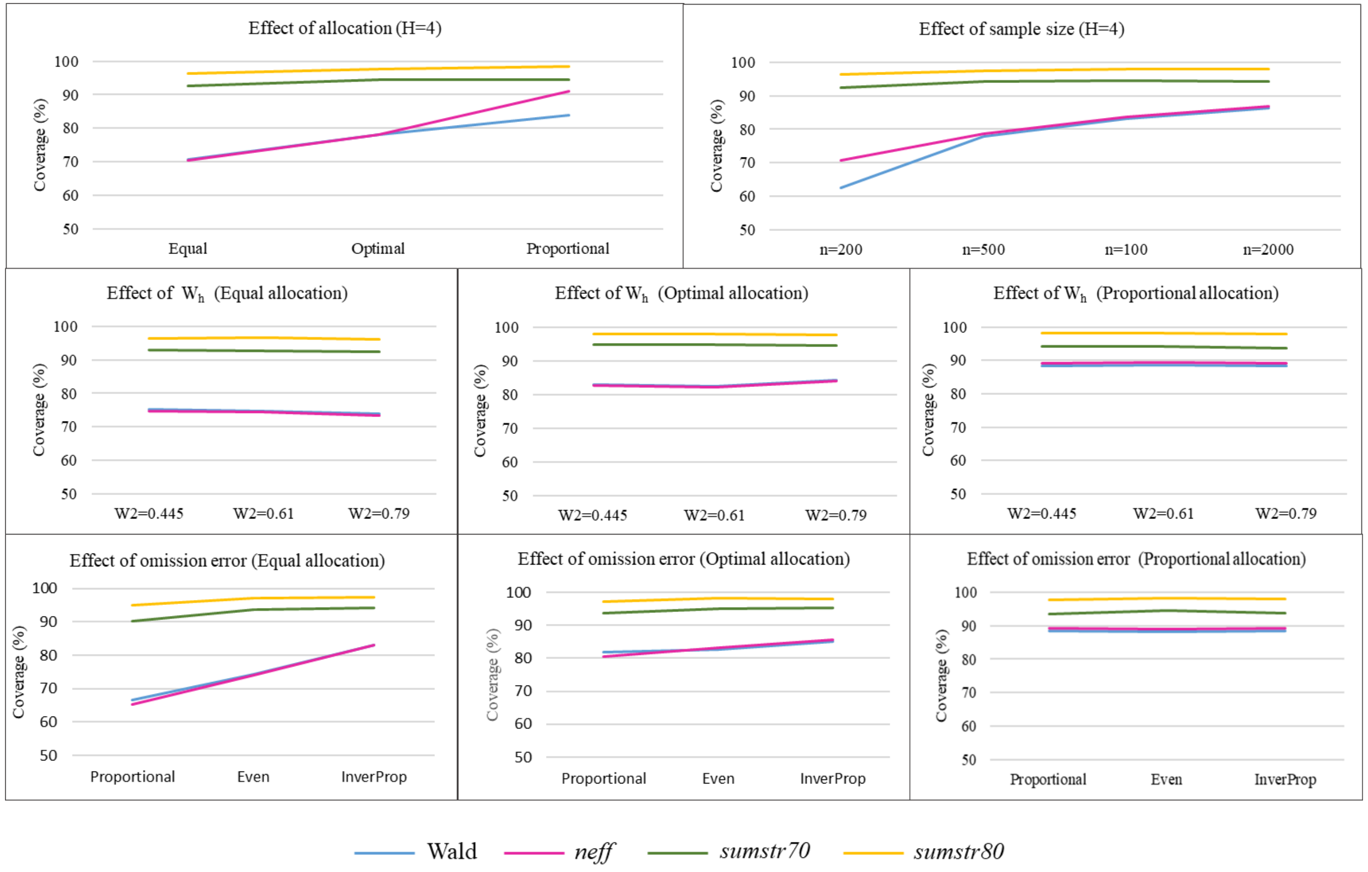


H=3

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

- ✓ Confidence Intervals
- ✓ Allocation
- ✓ Sample size
- ✓ Impact of stratum weight
- ✓ Impact of distribution of omission error



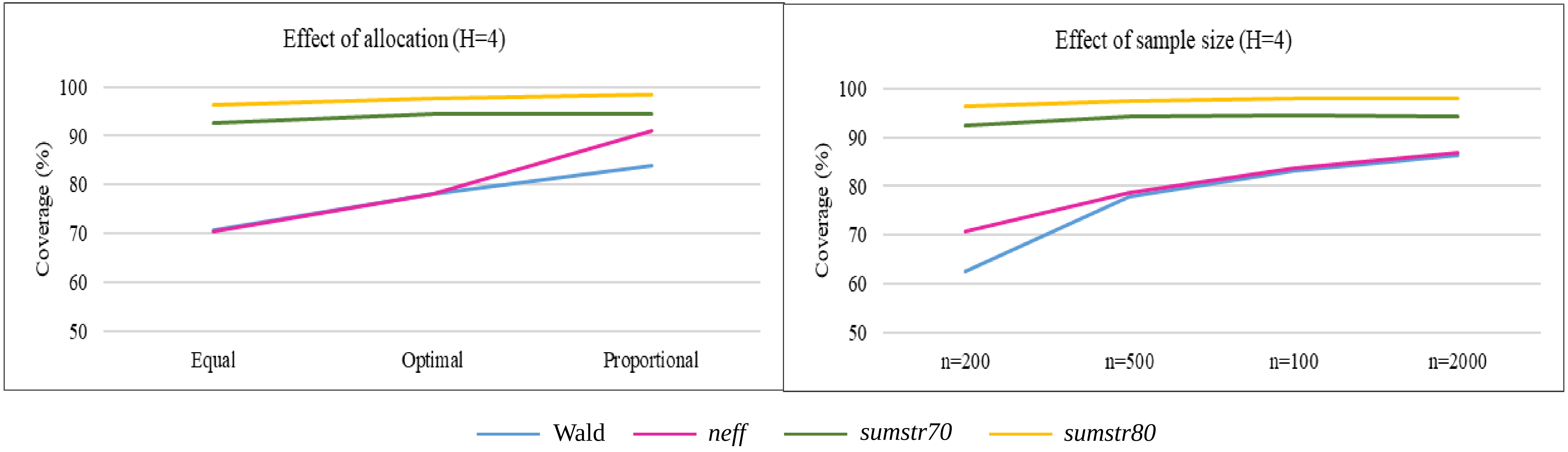
H=4

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

✓ Allocation

✓ Sample size

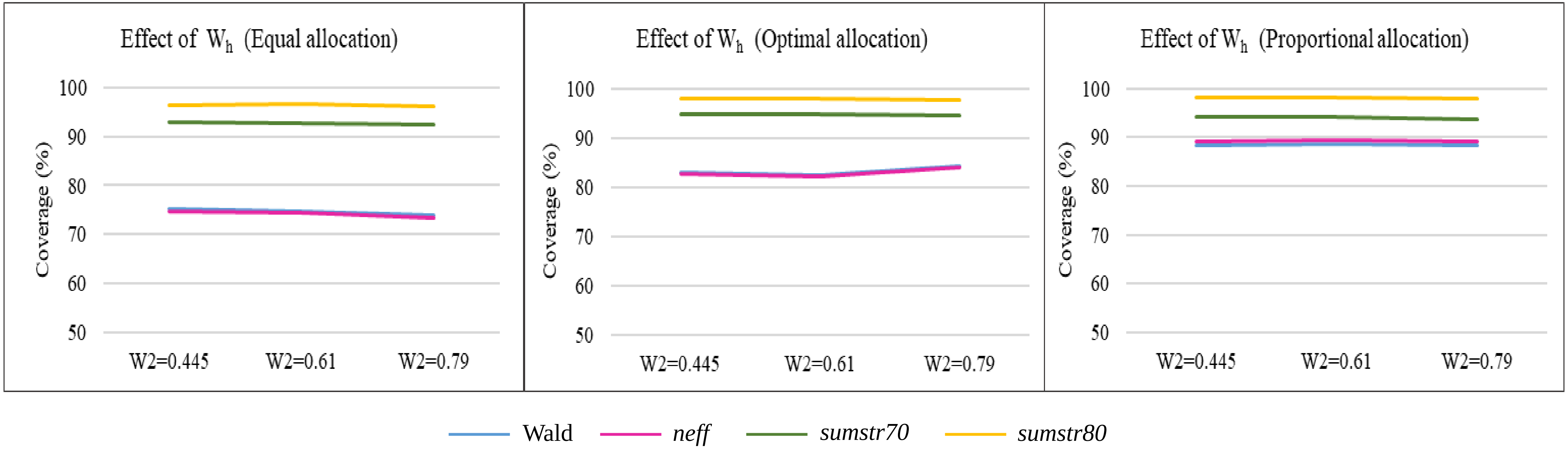


H=4

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

✓ Impact of stratum weight

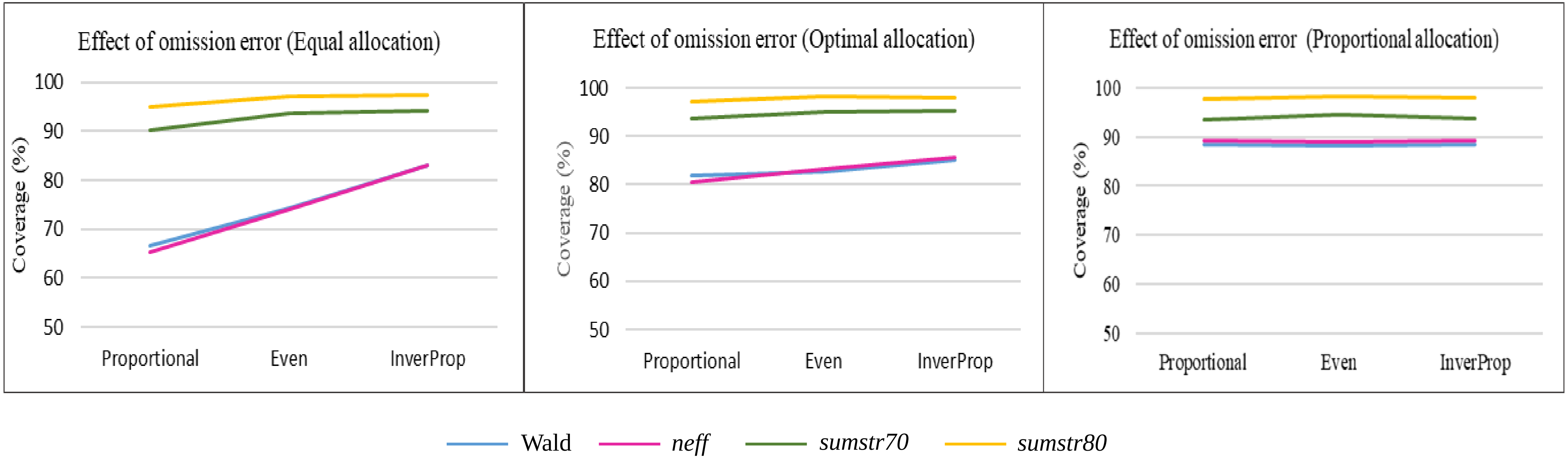


H=4

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

✓ Impact of distribution of area of omission error to strata



H=4

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Results

Case Study

CS #	#of Strata	p%	n	Wald	*neff*	*sumstr70*	*sumstr80*
CS1	3	0.09	1000	86	86	91	95
CS2	3	0.43	870	64	62	90	91
CS3	3	0.50	1000	89	90	88	93
CS4	3	6.16	994	89	89	91	98
CS5	4	6.77	997	89	90	96	99
CS6	4	9.70	4021	90	90	96	99
CS7	5	0.70	977	79	82	92	96
CS8	6	0.71	970	85	86	94	97
CS9	6	1.60	970	88	89	92	96
CS10	6	2.26	970	89	89	94	97
CS11	6	3.11	970	88	88	92	96
CS12	6	6.45	997	89	89	98	100
CS13	6	8.51	970	90	90	93	93

Actual sample size and sample allocation

	Equal				Optimal				Proportional			
CS #	Wald	*neff*	*sumstr70*	*sumstr80*	Wald	*neff*	*sumstr70*	*sumstr80*	Wald	*neff*	*sumstr70*	*sumstr80*
CS1	89	89	91	95	90	90	92	96	52	92	92	99
CS2	47	44	86	86	83	80	90	93	82	93	93	97
CS3	91	91	88	93	91	91	91	95	71	93	72	99
CS4	82	82	90	97	88	89	91	97	90	90	93	97
CS5	90	89	95	98	89	89	95	99	90	90	96	96
CS6	89	89	96	99	90	90	96	99	90	90	96	99
CS7	72	79	90	94	88	87	99	100	82	83	96	98
CS8	35	32	93	93	86	86	94	97	88	89	95	98
CS9	77	73	71	93	88	88	92	96	88	89	94	97
CS10	65	63	90	92	88	87	93	96	89	89	94	97
CS11	75	76	92	96	83	83	95	98	88	89	92	97
CS12	89	89	98	100	90	90	99	90	89	89	98	100
CS13	86	85	81	91	89	89	94	98	90	90	95	98

Populations and equal, optimal and proportional allocation

Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

❑ Discussion & Conclusion

- Wald interval was subject to substantial undercoverage, particularly for equal and optimal allocation when the sample size was small

- The *neff* approach performed well when proportional allocation was implemented.

- The *sumstrat* interval improved the coverage for equal and optimal allocation, AND coverage was stable over variation in weights, and omission error distribution.

```
if(n1[i]<2){
  i=i
}
else{

  SampleS1<-runif(n1[i], min=0,max=1)
  n11[i]<-length(which(SampleS1[]<=p1))
  Pro11[i]<- n11[i]/n1[i]

  SampleS2<-runif(n2[i], min=0,max=1)
  n21[i] <- length(which(SampleS2[]<=p2))
  Pro21[i]<- n21[i]/n2[i]

  SampleS3<-runif(n3[i], min=0,max=1)
  n31[i]<-length(which(SampleS3[]<=p3))
  Pro31[i]<- n31[i]/n3[i]

  VarP1 <- (Pro11[i]*(1-Pro11[i]))/(n1[i]-1)
  VarP2<- (Pro21[i]*(1-Pro21[i]))/(n2[i]-1)
  VarP3 <- (Pro31[i]*(1-Pro31[i]))/(n3[i]-1)

  Esti<-Pro11[i]*W1+Pro21[i]*W2+Pro31[i]*W3 #Estimate from a sample
  Var<-VarP1*W1*W1+VarP2*W2*W2+VarP3*W3*W3  #Estimate variance
  EstiProdAcc<- Pro11*W1/Esti  #Estimate prod's Acc
  EstiUserAcc<- Pro11  #Estimate User's Acc
  SE<-sqrt(Var) #Estimtae standard error

  #Wilson C.I
  WilsonCI_1<- BinomCI(n11[i], n1[i], conf.level = 0.85, method = "wilson")
  WilsonCI_2<- BinomCI(n21[i], n2[i], conf.level = 0.85, method = "wilson")
```

Developed the R program to help user to test out which interval is better to use

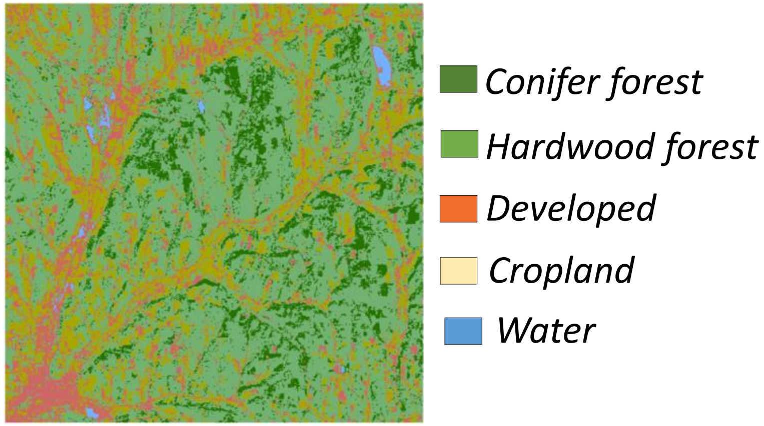
Manuscript 1 Confidence Intervals for a Proportion: Rare Classes

My Contribution

✓ Investigate the factors potentially related to coverage

- Omission error distribution?
- Map strata weights?
- Sample size ?
- Sample allocation ?
- Confidence Interval methods?

…… ?



✓ Expand the investigation of confidence interval properties for stratified sampling to H>2

H=3; H=4 or even more?

✓ Develop R programs to help users determine which confidence interval methods work well for a given application.

```r
if(n1[i]<2){
  i=i
}
else{

  SampleS1<-runif(n1[i], min=0,max=1)
  n11[i]<-length(which(SampleS1[]<=p1))
  Pro11[i]<- n11[i]/n1[i]

  SampleS2<-runif(n2[i], min=0,max=1)
  n21[i] <- length(which(SampleS2[]<=p2))
  Pro21[i]<- n21[i]/n2[i]

  SampleS3<-runif(n3[i], min=0,max=1)
  n31[i]<-length(which(SampleS3[]<=p3))
  Pro31[i]<- n31[i]/n3[i]

  VarP1 <- (Pro11[i]*(1-Pro11[i]))/(n1[i]-1)
  VarP2<- (Pro21[i]*(1-Pro21[i]))/(n2[i]-1)
  VarP3 <- (Pro31[i]*(1-Pro31[i]))/(n3[i]-1)

  Esti<-Pro11[i]*W1+Pro21[i]*W2+Pro31[i]*W3 #Estimate from a sample
  Var<-VarP1*W1*W1+VarP2*W2*W2+VarP3*W3*W3  #Estimate variance
  EstiProdAcc<- Pro11*W1/Esti  #Estimate prod's Acc
  EstiUserAcc<- Pro11  #Estimate User's Acc
  SE<-sqrt(Var) #Estimtae standard error

  #Wilson C.I
  WilsonCI_1<- BinomCI(n11[i], n1[i], conf.level = 0.85, method = "wilson")
  WilsonCI_2<- BinomCI(n21[i], n2[i], conf.level = 0.85, method = "wilson")
```

Manuscripts 2 & 3

Total Variance Estimation

Manuscript 2 & 3 Total Variance Estimation

❑ Introduction

Measurement Error

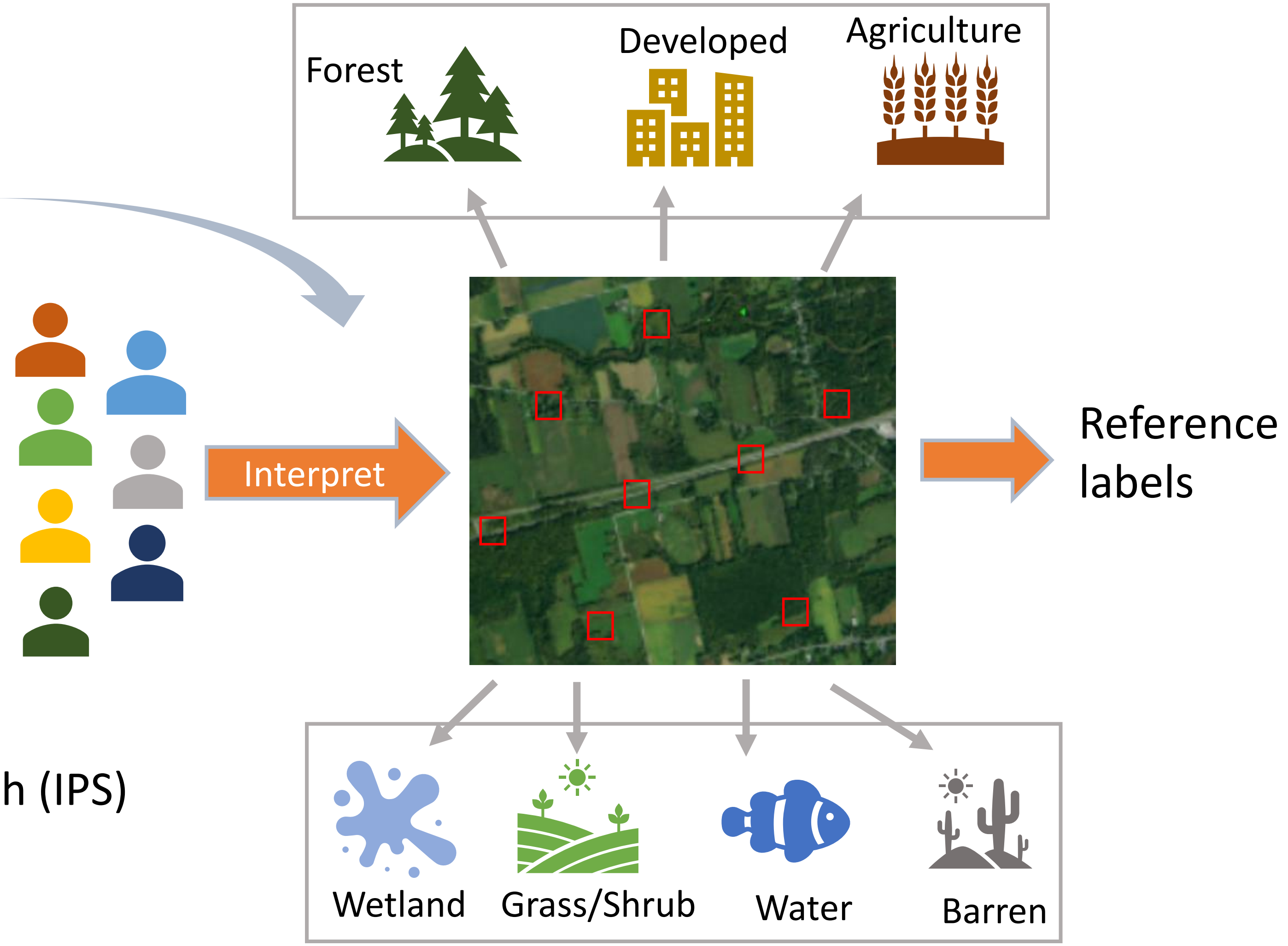
Sampling Variance

+

Measurement Variance

⇓

Total Variance



Total Variance Estimation

- Interpenetrating Subsampling Approach (IPS)
- Measurement Error Model (MEM)

Manuscript 2 & 3 Total Variance Estimation

- Data – LCMAP Sample

Land Cover, Monitoring, Assessment and Projection (LCMAP)

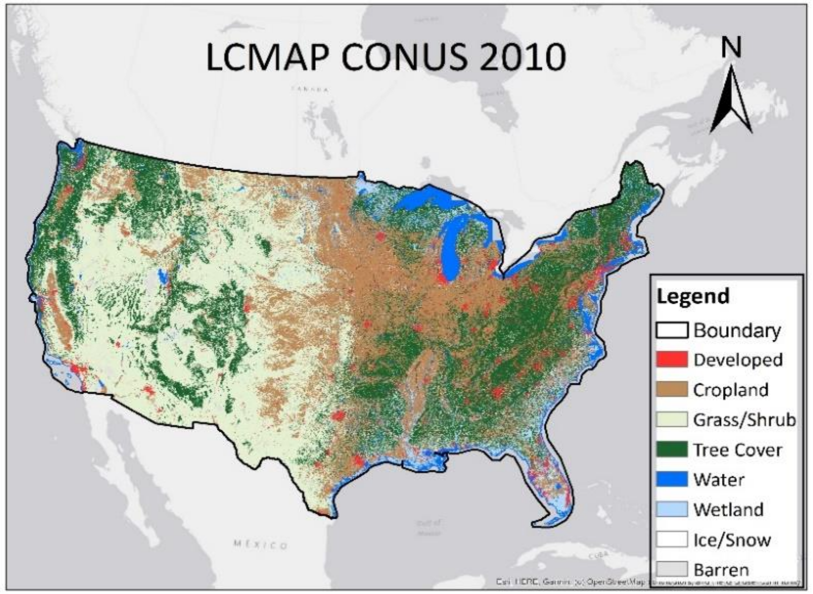
  - LCMAP Sample

Sample: n_s=25,000 in CONUS

Subsample: n_r=6080 duplicate interpretation

Interpreter 1

Interpreter 2



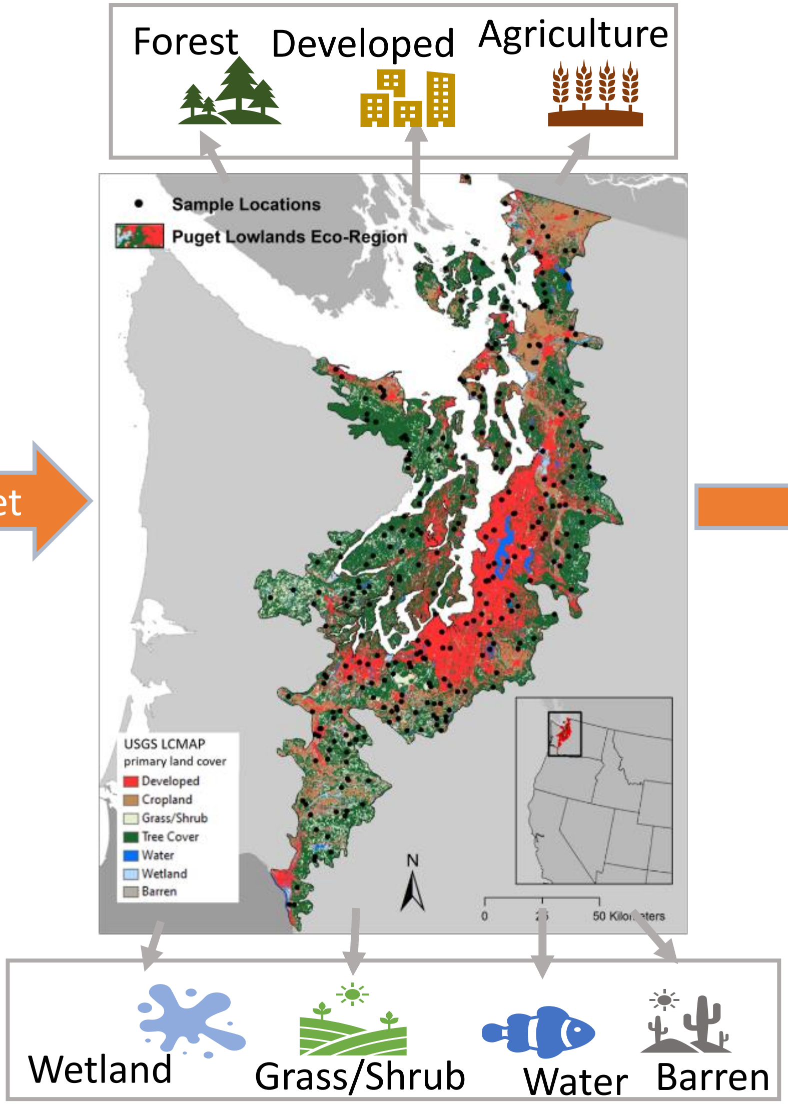
Class	Map% Sample	Map% Subsample	Interp 1 Subsample	Interp 2 Subsample
Tree Cover	25.8	25.6	28.2	28.1
Developed	3.9	3.9	5.3	5.3
Cropland	23.7	23.6	17.2	17.8
Grass/Shrub	35.2	35.2	38.3	37.6
Wetland	5.3	5.1	5.0	4.5
Water	5.0	5.4	5.2	5.5
Barren	1.1	1.0	0.9	1.2

Manuscript 2 & 3 Total Variance Estimation

- Data – LCMAP Pilot

7 Interpreters

Interpret



Reference labels

LCMAP Pilot Study
Puget Sound Region

- LCMAP Pilot

n_s=300 in Puget Sound Region

n_r=300 Interpreted by 7 interpreters

Source	Forest	Developed	Agriculture	Grass/Shrub	Wetland	Water	Barren	Rare_com
Map	51.7	26.0	13.7	3.0	3.0	2.3	0.0	8.3
Int 1	51.0	31.7	10.7	2.3	1.3	3.0	0.0	6.6
Int 2	54.0	29.3	9.7	2.0	2.3	2.3	0.3	6.9
Int 3	45.7	28.3	8.0	12.0	3.3	2.7	0.0	18
Int 4	51.7	30.3	8.3	2.7	5.0	2.0	0.0	9.7
Int 5	53.0	25.3	10.3	6.3	2.7	2.3	0.0	11.3
Int 6	47.3	30.7	10.0	5.7	3.3	2.7	0.3	12.0
Int 7	48.7	29.7	9.3	5.0	5.0	2.3	0.0	12.3
Majority	51.0	28.7	9.0	3.0	3.7	2.3	0.0	9.3

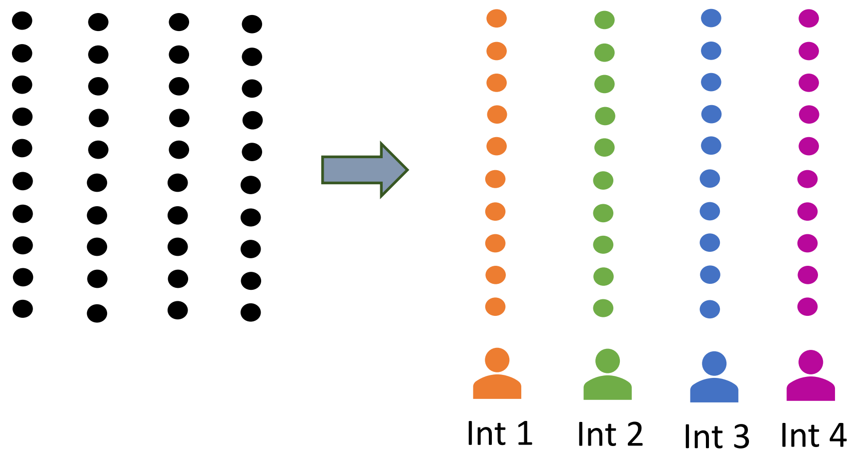
Manuscript 2 & 3 Total Variance Estimation

❑ Methods

✓ Interpenetrating Subsampling Approach (IPS)

Sample pixels

Divide sample pixels into g groups



g interpreters

$$\hat{V}_{Total} = \frac{N^2}{g(g-1)} \sum_{i=1}^{g} (\bar{y}_{k_i} - \bar{y}_k)^2$$

Estimator of total variance

$$MS_b = \frac{m}{(g-1)} \sum_{i=1}^{g} (\bar{y}_{k_i} - \bar{y}_k)^2$$

$$MS_w = \frac{1}{g(m-1)} \sum_{i=1}^{g} \sum_{k_i} (y_k - \bar{y}_{s_i})^2$$

$$\hat{V}_{12} = N^2 \frac{m-1}{m} \frac{(MS_b - MS_w)}{n}$$

Correlated measurement variance

$$\hat{\rho}_{int} = \frac{\frac{(MS_b - MS_w)}{m}}{\left(\frac{MS_b - MS_w}{m}\right) + MS_w}$$

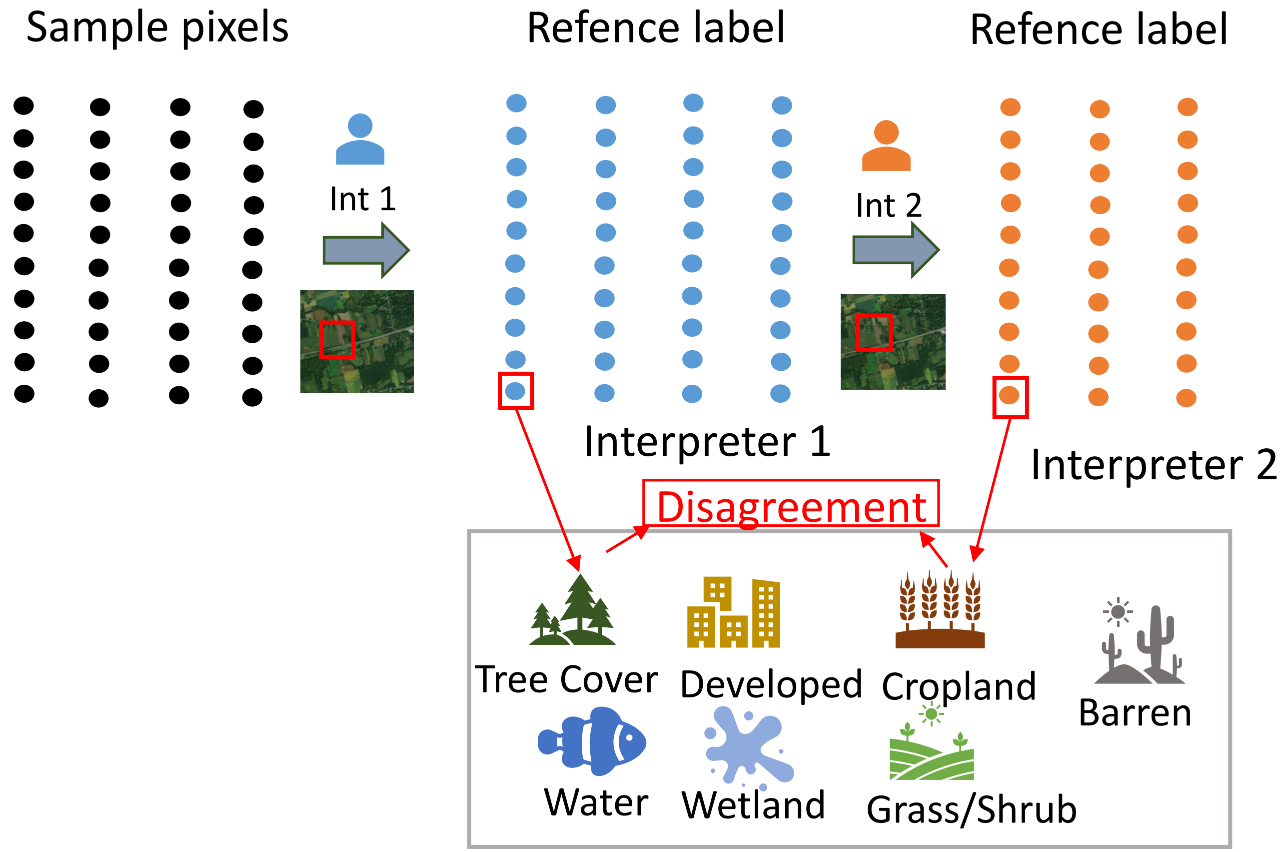
Coefficient of intraclass correlation

An advantage of this method is that it does not require the additional cost of duplicate interpretations. But it does require a more complex, structured assignment of sample units to interpreters.

Manuscript 2 & 3 Total Variance Estimation

❑ Methods

✓ Repeated Measurement Error Model (MEM)



Variance unaffected by sampling

Sampling variance

$$\hat{V}_{total1} = \hat{V}_{1cen} + \hat{V}_{stand}$$

MEM_1

$$\hat{V}_{total2} = \hat{V}_1 + \hat{V}_2 = \hat{V}_{11} + \hat{V}_{12} + \hat{V}_2$$

$$\approx \hat{V}_{11} + \hat{V}_{12} + \hat{V}_{stand}$$

MEM_2

Simple measurement variance

Correlated measurement variance

The MEM can be used to estimate total variance as well as the variance components contributing to total variance, but additional cost to obtain the repeat measurements is required in the approach.

Manuscript 2 & 3 Total Variance Estimation

Methods

✓ Repeated Measurement Error Model (MEM)

$$\hat{V}_{total1} = \hat{V}_{1cen} + \hat{V}_{stand}$$

MEM_1

$$\hat{V}_{1cen} = \frac{n_s}{2n_r}\sum_r \frac{z_k^2}{\pi_k} + \frac{n_s(n_s-1)}{2n_r(n_r-1)}\sum_{k\neq l}\sum_r z_k z_l \Big/ \pi_{kl}$$

$$\hat{V}_{stand} = (1 - n/N)\frac{\hat{p}(1-\hat{p})}{(n-1)}$$

Simple random sampling

$$\hat{V}_{stand} = \sum_{h=1}^{H} W_h^2 (1 - \frac{n_h}{N_h})\hat{p}_h(1-\hat{p}_h)/(n_h - 1)$$

Stratified random sampling

$$\hat{V}_{total2} = \hat{V}_1 + \hat{V}_2 = \hat{V}_{11} + \hat{V}_{12} + \hat{V}_2$$

MEM_2

$$\approx \hat{V}_{11} + \hat{V}_{12} + \hat{V}_{stand}$$

$$\hat{V}_{11} = \frac{n_s}{2n_r}\sum_r (z_k/\pi_k)^2$$

$$\hat{V}_{12} = \frac{n_s(n_s-1)}{2n_r(n_r-1)}\left\{\left(\sum_r z_k \Big/ \pi_k\right)^2 - \sum_r (z_k/\pi_k)^2\right\}$$

V_{stand} is a biased estimate of V_2 (overestimate) so it would generally fall between estimating V_2 and V_{total}

Manuscript 2 & 3 Total Variance Estimation

❑ Methods

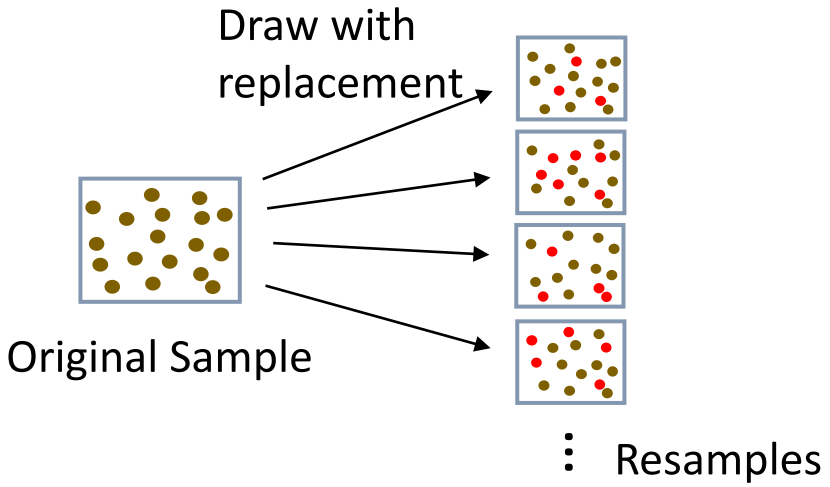
- ✓ Monte-Carlo Hybrid Variance Estimator (MCHybrid)
  - McRoberts et al. (2018)

Contents lists available at ScienceDirect
ISPRS Journal of Photogrammetry and Remote Sensing
journal homepage: www.elsevier.com/locate/isprsjprs

The effects of imperfect reference data on remote sensing-assisted estimators of land cover class proportions

Ronald E. McRoberts[a,*], Stephen V. Stehman[b], Greg C. Liknes[a], Erik Næsset[c], Christophe Sannier[d], Brian F. Walters[a]
[a] Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA
[b] Forest and Natural Resources Management, State University of New York, Syracuse, NY, USA
[c] Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway
[d] Systèmes d'Information à Référence Spatiale (SIRS), Villeneuve d'Ascq, France

- ✓ Monte-Carlo Measurement Error Model (MCMEM)
  - Bootstrapping
  - Developed in this study



- ✓ Descriptive Measures of Interpreter Variability
  - Cochran (1977)

$$\left\{ \begin{array}{l} \text{Response variance } \hat{\sigma}_d^2 \\ \text{Index of inconsistency } \hat{I} \end{array} \right.$$

Manuscript 2 & 3 Total Variance Estimation

❑ Results

Interpreter inconsistency ($\hat{I}$%) for the LCMAP sample and LCMAP Pilot sample

	Index $\hat{I}$	
	LCMAP	LCMAP
Class	sample	Pilot
Water	5	7
Cropland	11	22
TreeCover	12	15
Developed	17	15
Grass/Shrub	17	65
Wetland	21	47
Barren	49	--

- Water Tree cover and developed had the smallest index $\hat{I}$ for the two samples.
- Grass/shrub and Wetland had the largest interpreter inconsistency in both LCMAP sample data and LCMAP Pilot data.

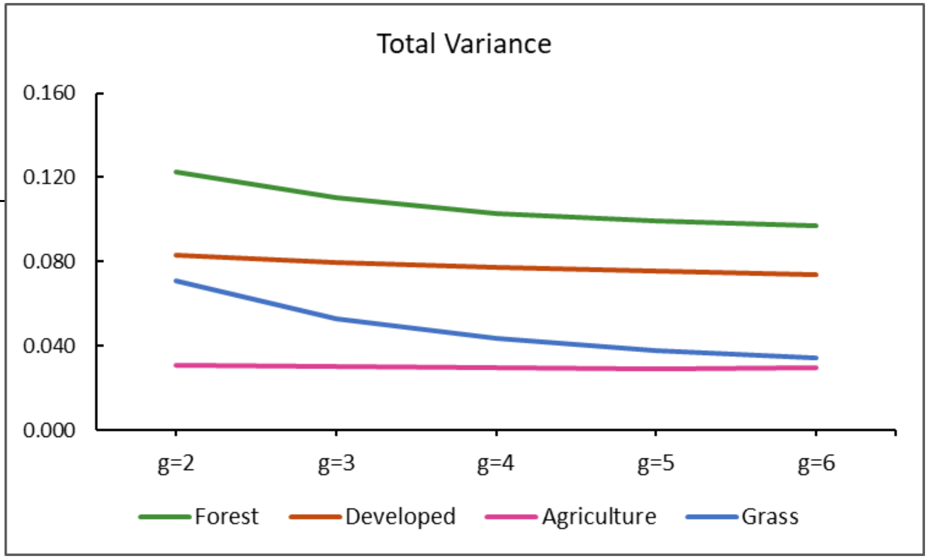
Manuscript 2 & 3 Total Variance Estimation

❑ Results ✓ Interpenetrating Subsampling Approach (IPS)

LCMAP Pilot data

	Class	g=2	g=3	g=4	g=5	g=6	$\hat{V}_{stand}$ (%)
$\hat{V}_{total}$ (%)	Forest	0.123	0.110	0.103	0.099	0.097	0.084
	Developed	0.083	0.079	0.077	0.075	0.074	0.068
	Agriculture	0.031	0.030	0.030	0.029	0.030	0.027
	Grass/Shrub	0.071	0.053	0.044	0.038	0.034	0.010
$\hat{V}_{total}/\hat{V}_{stand}$	Forest	1.46	1.31	1.22	1.18	1.15	
	Developed	1.23	1.17	1.13	1.11	1.08	
	Agriculture	1.13	1.11	1.09	1.08	1.10	
	Grass/Shrub	7.06	5.26	4.37	3.80	3.44	

Simple Random Sampling



Manuscript 2 & 3 Total Variance Estimation

- Results
  - ✓ Interpenetrating Subsampling Approach (IPS)

LCMAP Pilot data

	Class	g=2	g=3	$\hat{V}_{stand}(\%)$
$\hat{V}_{total}(\%)$	Forest	0.069	0.055	0.025
	Developed	0.047	0.042	0.028
	Agriculture	0.016	0.015	0.013
	Rare_com	0.092	0.069	0.019
$\hat{V}_{total}/\hat{V}_{stand}$	Forest	2.76	2.20	
	Developed	1.68	1.50	
	Agriculture	1.23	1.15	
	Rare_com	4.84	3.63	

Stratified Random Sampling

Manuscript 2 & 3 Total Variance Estimation

❑ Results ✓ Repeated Measurement Error Model (MEM)

LCMAP Sample data

Class	Design	$\hat{p}\%$	$\sqrt{\hat{V}_{11}}$	$\sqrt{\hat{V}_{12}}$	$\sqrt{\hat{V}_{1}}$	$\sqrt{\hat{V}_{1cen}}$	SE_{stand}	SE_{MEM_1}	SE_{MEM_2}
TreeCover	SRS	28.2	0.10	(0.18)	(0.15)	(0.18)	**0.28**	**0.22**	**0.24**
Developed	SRS	5.3	0.06	(0.12)	(0.10)	(0.12)	**0.14**	**0.08**	**0.10**
Cropland	SRS	17.2	0.08	0.04	0.09	0.04	**0.24**	**0.24**	**0.26**
Grass/Shrb	SRS	38.3	0.13	(0.25)	(0.22)	(0.25)	**0.31**	**0.17**	**0.21**
Wetland	SRS	5.0	0.06	0.28	0.28	0.28	**0.31**	**0.41**	**0.42**
Water	SRS	5.2	0.03	0.05	0.06	0.05	**0.31**	**0.31**	**0.31**
Barren	SRS	0.9	0.05	(0.10)	(0.08)	(0.10)	**0.31**	**0.29**	**0.30**
TreeCover	STR	28.2	0.10	(0.10)	(0.04)	(0.18)	**0.16**	**(0.07)**	**0.16**
Developed	STR	5.4	0.06	(0.10)	(0.08)	(0.12)	**0.10**	**(0.06)**	**0.06**
Cropland	STR	17.1	0.08	0.12	0.15	0.03	**0.16**	**0.16**	**0.21**
Grass/Shrb	STR	38.3	0.13	(0.25)	(0.22)	(0.25)	**0.20**	**(0.15)**	**(0.07)**
Wetland	STR	4.9	0.08	0.35	0.36	0.27	**0.09**	**0.28**	**0.37**
Water	STR	5.3	0.04	0.11	0.12	0.05	**0.05**	**0.07**	**0.13**
Barren	STR	0.9	0.05	(0.08)	(0.06)	(0.10)	**0.05**	**(0.08)**	**(0.04)**

Total variance estimated from MEM_1 and MEM_2 was negative more often in STR.

MEM_2 resolved the negative variance problem of MEM_1 to some extent.

Manuscript 2 & 3 Total Variance Estimation

❑ Results

LCMAP Pilot data

Comparison of the Different Total Variance Estimators

➢ Simple random sampling

Class	SE_{stand}	MEM_1	MEM_2	MCMEM_D
TreeCover	2.89	3.50 (1.09)	3.59 (1.07)	2.88
Developed	2.62	2.83 (0.75)	2.99 (0.73)	2.63
Cropland	1.66	1.68 (0.26)	1.88 (0.24)	1.69
Grass/Shrub	0.99	2.37(2.07)	2.62 (2.00)	1.28
Wetland	1.08	1.52 (0.58)	1.65 (0.54)	1.04
Water	0.87	0.90 (0.06)	0.93 (0.07)	0.90

- The standard errors obtained from standard variance estimator were close to the standard errors from MCMEM method using majority labels among 7 different interpreters.
- Two MEM methods show higher values of SEs compared to other methods

- The largest difference of SEs between MEM and other methods was observed in grass/shrub, this class was also observed with larger inconsistency between 21 pairs of 7 interpreters

Manuscript 2 & 3 Total Variance Estimation

❑ Results

LCMAP Pilot data

Comparison of the Different Total Variance Estimators

➢ Stratified random sampling

Class	SE_{stand}	MEM_1	MEM_2	MCMEM_D	MCHyb
Tree Cover	1.57	2.64 (1.37)	2.49 (0.86)	1.73	2.27 (0.05)
Developed	1.68	2.06 (0.94)	2.07 (0.53)	1.79	2.44 (0.02)
Cropland	1.15	1.21 (0.37)	1.47 (0.19)	1.19	1.58 (0.01)
Grass/Shrub	0.98	2.38 (2.08)	2.18 (1.85)	1.26	1.24 (0.20)

- In general, the standard error using different total variance estimators are higher than the SE of standard variance, indicating measurement variance is an important contribution in total variance, in practice, we need to incorporate it in total variance estimation.
- Hybrid estimator has relatively “stable” SE than MEM methods among all 21 pairs of interpreters – with smaller values of standard deviation among all 21 pairs
- Same as in simple random sampling, the SEs from two MEM approaches and hybrid estimator were relatively higher than other methods.

Manuscript 2 & 3 Total Variance Estimation

❑ Conclusion & Discussion

Manuscript 2. Using Interpenetrating Subsampling to Incorporate Interpreter Variability into Estimation of the Total Variance of Land Cover Area Estimates

- Interpenetrating subsampling provides an easy way to estimate the total variance, and it is a practical approach for large-scale studies.
- As the number of subgroups increased, the total variance estimated by IPS decreased.
- A greater number of subgroups in IPS resulted in smaller variability of the estimated total variance over different random partitions of the sample into the IPS subgroups.

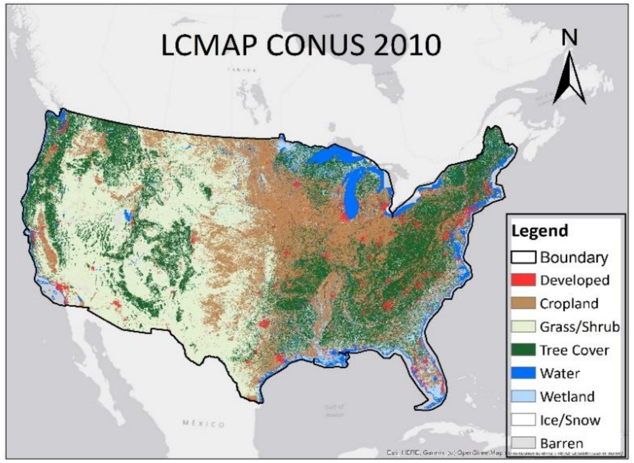
Manuscript 2 & 3 Total Variance Estimation

❑ Conclusion & Discussion

Manuscript 3. Applications of Total Variance Estimation Incorporating Reference Data Variability to Area Estimates of Land Cover

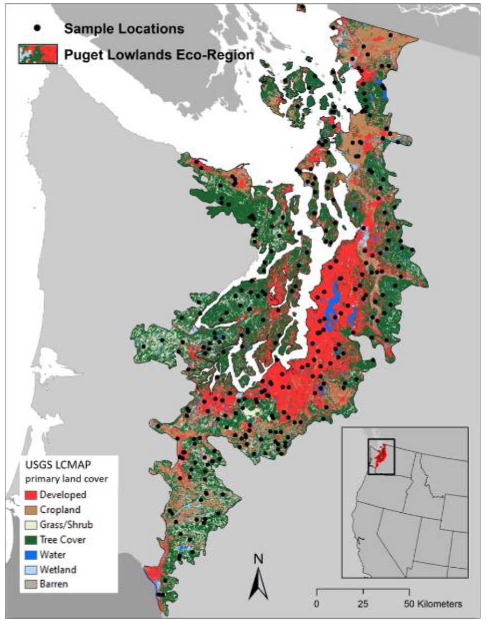
- Compare different total variance estimators applied to land cover reference data.
- Derived the formulas of the MEM estimator for stratified sampling.
- This study examined the MEM applied to both simple random and stratified random estimates at the country level dataset—LCMAP CONUS.
- Negative variance estimates were observed for MEM methods.
- When duplicated labeling of all sample pixels was possible, two Monte-Carlo simulation approaches – MCHybrid and MCMEM provide alternative reliable ways to estimate total variance.

Manuscript 2 & 3 Total Variance Estimation



My Contribution

- ✓ First study which introducing IPS to the environmental field
- ✓ Explore the properties of IPS
  - Variability over different random partitions into subgroups
  - Variability over different pairs of interpreters
- ✓ Provide the recommendations to reduce the contribution of interpreter variance
  - Interpreter variance can contribute substantially to the total variance
  - More interpreters help to reduce the total variance in IPS
- ✓ Apply different total variance estimators to operational large-area land cover monitoring program-- LCMAP
  - Contribute the formulas of the MEM estimator for stratified sampling
  - Problem of negative estimates of variance estimates



Dissertation

❑ Overall

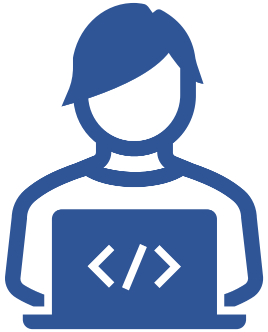
(1) Extend the evaluation of confidence interval methods identifying factors of the population and sampling design that are related to coverage to applications in which more than two strata are present in stratified random sampling.

(2) Apply the IPS approach to incorporate interpreter variability into estimation of the total variance under SRS and STR, and determine the impact of the number of interpreters (subgroups) and random partition of sample to subgroups on estimate.

(3) Extend estimation of total variance incorporating reference data variability to area estimates in an operational land cover monitoring program by using the MEM and Monte-Carlo simulation estimators, and compare the variance estimates resulting from the different approaches.

Future Work

❑ Future Work



- ➢ Technology and complex simulation program
  - More comprehensive simulation calculations
  - Provide further detailed recommendations on the use of confidence intervals under complex sampling designs.

- ➢ Extension of the models estimating total variance
  - Correlation between sampling and response variance
  - Extending to continuous variables
  - Fellegi (1964)

Journal of the American Statistical Association

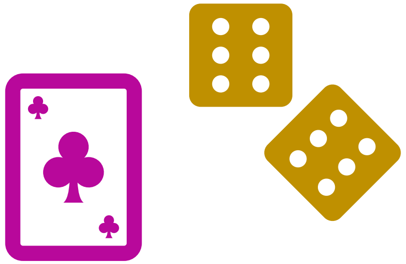
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uasa20

Response Variance and its Estimation

I. P. Fellegi [a]

[a] Dominion Bureau of Statistics , Canada

- ➢ Evaluating confidence interval performance taking into account interpreter variability
  - Use SE from total variance estimator which incorporates the interpreter variability when estimating confidence interval

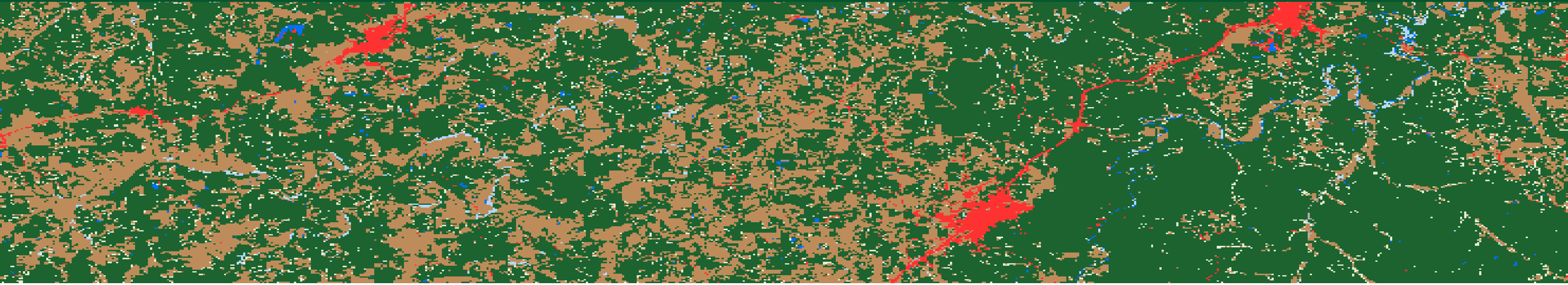


Acknowledgement

I want to take this opportunity to express my deepest sincerest gratitude to my advisor Dr. Stephen Stehman, who provide me strong support, valuable advice and persistent encouragement through out my academic journey. I would like to thank my committees Dr. Diane Kiernan and Dr. Eddie Bevilacqua for their kindness, understanding and efficient instruction in my research and teaching assistant work. I also warmly thank my examiners Dr. Alexandra Yurievna Tyukavina (University of Maryland), Dr. James Wickham (U.S. Environmental Protection Agency) and the chair Prof. Matthew Smith for their strong support and help.

Special thanks to my dear friends Lin Wu, Yun Zhao and Liyuan Feng for their great encouragement, to Anusha, Basanta, Fern, and Abishek for the kind help. And sincerely thank all of you attending this capstone.

This work was supported with funding from the U.S. Geological Survey (USGS), Graduate Program in Environmental Science (GPES) and department of Sustainable Resources Management (SRM) of ESF.



Thank you

Questions?

Dingfan Xing
dixing@esf.edu
SUNY-ESF
Oct 23, 2023