The Lies in LiDAR

2016 NEMO/SOMENS joint meeting, Concord, MA

John Kershaw, Aaron Weiskittel, Mark Ducey

November 13 – 15, 2016

Best Practices?

The List of Lies

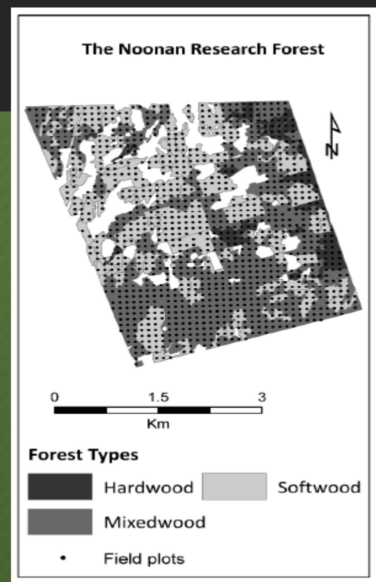
1. Ground sampling design is not important as long as you measure what you are interested in
2. LiDAR is a physical measurement, so that makes it different from other remotely sensed data
3. LiDAR metrics actually mean something and are not scale-dependent
4. Georeferencing of plots
5. Fixed area plots are superior to variable probability plots
6. Plot size = LiDAR extraction size = Estimation size
7. LiDAR data is LiDAR data (not all LIDAR sensors are created equally)
8. Leaf-off or leaf-on is superior
9. LiDAR “sees” individual trees
10. xyz points are useful for forest health and species classification
11. LiDAR is a physical measurement, so it has to match field-measured height to be useful.
12. Why exactly should we be excited to discover that height is well correlated with volume/biomass?
13. randomForest as a multivariate analytical tool (inconsistent estimates across parameters)
14. randomForest (or some other multivariate technique) can predict my variable (e.g. downed wood volume) from LiDAR, so LiDAR is measuring my variable
15. LiDAR is modern scientific forest management (somehow it’s not a static measurement, dated as soon as it is acquired)
16. Error free inventory

The List of Lies

1. Ground sampling design is not important as long as you measure what you are interested in
2. LiDAR is a physical measurement, so that makes it different from other remotely sensed data
3. LiDAR metrics actually mean something and are not scale-dependent
4. Georeferencing of plots
5. **Fixed area plots are superior to variable probability plots**
6. **Plot size = LiDAR extraction size = Estimation size**
7. LiDAR data is LiDAR data (not all LIDAR sensors are created equally)
8. Leaf-off or leaf-on is superior
9. LiDAR “sees” individual trees
10. xyz points are useful for forest health and species classification
11. LiDAR is a physical measurement, so it has to match field-measured height to be useful.
12. Why exactly should we be excited to discover that height is well correlated with volume/biomass?
13. **randomForest as a multivariate analytical tool (inconsistent estimates across parameters)**
14. randomForest (or some other multivariate technique) can predict my variable (e.g. downed wood volume) from LiDAR, so LiDAR is measuring my variable
15. LiDAR is modern scientific forest management (somehow it’s not a static measurement, dated as soon as it is acquired)
16. Error free inventory

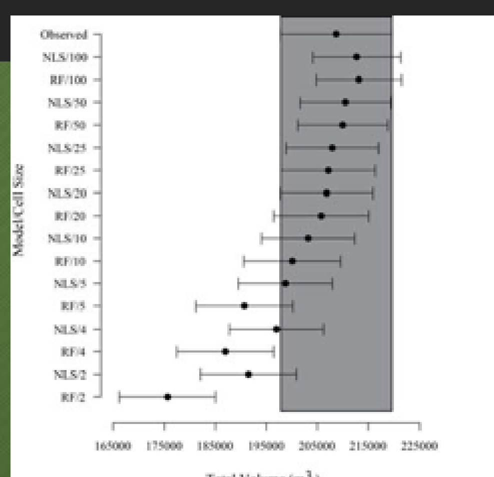
Noonan Research Forest LiDAR Study

- 1500 ha
- 1480 2M big BAF horizontal point samples (100m grid)
- 105 0.04ha fixed area CFI plots
- three 50m × 50m mapped plots
- 25ha 10m resolution tree tally
- Leaf-off LiDAR (~6 returns/m^2)



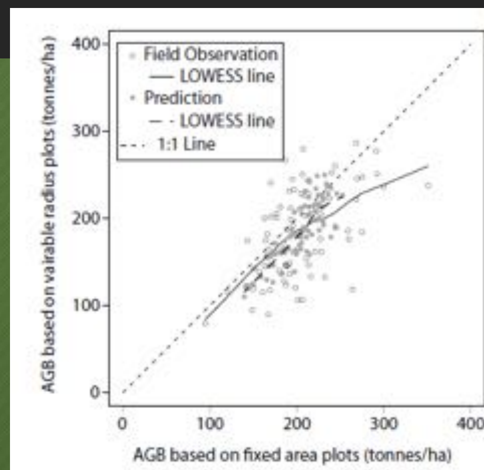
6) Plot size = LiDAR extraction size = Estimation size

- Models developed at 10 m cell resolution
- Wall-to-wall estimates made at resolutions of 2m - 100m
- Across a wide range of cell sizes no significant difference from the ground estimates
- NLS more robust than RF to scale changes
- Max Canopy Height hadthe most influence



5) Fixed area plots are superior to variable probability pointss

- 2M BAF versus 0.04ha circular plots
- No differences in predictions using NLS-based LiDAR models



13) randomForest as a multivariate analytical tool

- Goodness-of-fit Local Data

Data Source		Fit	NLME					Random Forest				
Model	Test	Statistic	10	15	20	25	30	10	15	20	25	30
NRF	NRF	R^2	0.68	0.67	0.64	0.63	0.61	0.85	0.84	0.83	0.82	0.82
		RMSE	33.4	34.4	35.5	36.4	37	22.8	23.5	24.3	24.9	25.5
		Mean Bias	-0.3	-0.3	-0.2	-0.2	-0.1	-0.11	-0.07	-0.05	-0.01	-0.1
		Abs Error	25.9	26.8	27.8	28.5	29.1	17.7	18.2	19.1	19.6	20
PEF	PEF	R^2	0.43	0.53	0.54	0.54	0.53	0.8	0.84	0.83	0.84	0.84
		RMSE	45.4	41	40.6	40.8	41.1	26.7	24.3	24.7	23.8	24.3
		Mean Bias	0.33	-0.2	-0.5	-0.5	-0.6	-0.07	0.16	0.08	0.15	0.34
		Abs Error	35.5	32.2	31.8	31.8	31.7	18.7	16.6	17.3	16.7	17.1

13) randomForest as a multivariate analytical tool

- Equivalence of Prediction Local Data

Equivalence Test			LiDAR Extraction Radius				
			10	15	20	25	30
Field	NLME{N/N+P/P}	H{0}	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	-0.26	-0.28	-0.25	-0.21	-0.18
		Std Dev	34.5	34.95	35.92	36.74	37.37
		Region of Sim	15%	10%	10%	10%	10%
	RF{N/N+P/P}	H{0}	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	-0.10	-0.05	-0.04	0.01	-0.07
		Std Dev	23.09	23.57	24.36	24.83	23.58
		Region of Sim	15%	15%	15%	15%	15%
NLME(N/N+P/P)	RF{N/N+P/P}	H{0}	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	0.16	0.22	0.21	0.21	0.11
		Std Dev	14.96	14.84	14.87	15.18	15.41
		Region of Sim	25%	25%	25%	20%	20%

13) randomForest as a multivariate analytical tool

- Goodness-of-fit Non-Local Data

Data Source		Fit	NLME						Random Forest				
Model	Test	Statistic	10	15	20	25	30		10	15	20	25	30
NRF	PEF	R^2	0.4	0.48	0.48	0.47	0.46		0.35	0.37	0.39	0.39	0.39
		RMSE	46.3	43.2	43.1	43.6	44.1		48.3	47.5	46.7	46.6	46.7
		Mean Bias	3.56	5.36	7.12	6.93	7.84		14.4	14.51	15.03	14.52	14.36
		Abs Error	36.2	32.7	32.4	32.9	32.9		36	35.4	35.4	35.2	35.2
PEF	NRF	R^2	0.64	0.58	0.59	0.58	0.58		0.62	0.57	0.53	0.51	0.49
		RMSE	35.8	38.7	38.1	38.5	38.5		36.6	38.9	40.7	41.8	42.5
		Mean Bias	5.42	11.7	8.2	7.42	7.42		-5.39	-11.9	-14.1	-15.4	-15.5
		Abs Error	27.8	29.7	29.4	29.8	29.8		28.9	31.4	32.7	33.5	34

13) randomForest as a multivariate analytical tool

- Equivalence of Prediction Non-Local Data

Equivalence Test			LiDAR Extraction Radius				
			10	15	20	25	30
Field	NLME{N/P+P/N}	H{0}	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	5.27	11.21	8.12	7.38	6.45
		Std Dev	36.34	37.41	37.66	38.23	38.72
		Region of Sim	25%	40%	30%	30%	25%
	RF{N/P+P/N}	H{0}	Reject	Reject	Reject	Reject	Reject
		Mean Diff.	-3.87	-9.86	-11.89	-13.01	-13.24
		Std Dev	37.41	38.43	39.44	40.07	40.78
		Region of Sim	20%	35%	40%	45%	45%
NLME{N/N+P/P}	NLME{N/P+P/N}	H{0}	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	5.54	11.06	8.37	7.59	6.62
		Std Dev	11.06	11.06	10.91	10.08	9.66
		Region of Sim	50%	≤50%	≤50%	≤50%	≤50%
	RF{N/P+P/N}	H{0}	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-3.60	-9.59	-11.63	-12.89	-13.07
		Std Dev	13.34	17.99	17.99	17.92	18.15
		Region of Sim	40%	≤50%	≤50%	≤50%	≤50%
NLME{N/P+P/N}	RF{N/N+P/P}	H{0}	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-5.38	-11.27	-8.16	-7.38	-6.51
		Std Dev	17.94	19.31	18.24	18.21	18.14
		Region of Sim	45%	≤50%	≤50%	≤50%	≤50%
	RF{N/P+P/N}	H{0}	Not rej.	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-9.14	-21.07	-20.00	-20.48	-19.69
		Std Dev	18.04	21.42	20.78	20.22	19.45
		Region of Sim	≤50%	≤50%	≤50%	≤50%	≤50%
RF{N/N+P/P}	RF{N/P+P/N}	H{0}	Reject	Not rej.	Not rej.	Not rej.	Not rej.
		Mean Diff.	-3.77	-9.81	-11.85	-13.11	-13.17
		Std Dev	19.45	19.84	20.51	20.59	21.24
		Region of Sim	≤50%	≤50%	≤50%	≤50%	≤50%

A return to some first principles?

- A robust, well-designed ground sample will get you a long way
- You need a sample to estimate a total
- LiDAR helps you distribute that total over your forest

- What good is your model if you can’t generalize or explain it?