Estimating Saturated Hydraulic Conductivity using Decision Tree Analysis

A. Shirley, D.E. Radcliffe, A. Nemes

Dept. of Crop and Soil Sciences, University of Georgia, Athens, GA

aubshi@gmail.com

Abstract

Pedotransfer functions (PTFs) are useful tools that can help predict saturated hydraulic conductivity (Ksat). However, at present, most PTFs only utilize soil texture and bulk density. There are numerous other properties, including soil structure, which are available in soil survey databases that may be useful for more accurately predicting Ksat. In our study we want to determine which of these, if any, will give a more accurate prediction of Ksat. We have used soil profile descriptions from the S-124 regional project dataset. This is a Southern Cooperative Bulletin Series which contains 21 soil series descriptions from the Southeastern United States. These bulletins contain qualitative soil structure descriptions. We have broken down the descriptors into four qualitative and two quantitative groups: horizon notation, texture class, ped size, crack orientation, bulk density, and particle size distribution. Qualitative variables are represented by zeros and ones, determined by the sample's membership in the respective groups. Preliminary runs show reasonable estimations for Ksat using the decision-tree model in our study. However, the sand texture class prediction was lower than expected with a value of 13.21 cm/day. The overall root mean squared residual (RMSR) was 0.3911 for Log (Ksat), thus indicating a good prediction of Ksat when compared to values measured in the field.

Methods

- Followed the binary procedure for qualitative variables set forth by Lilly et al. (2008)
- Qualitative variables included: structure (ped size, crack orientation, and textural class)
- Members received ones or zeroes (members/non-members)
- Quantitative variables were input directly (bulk density, organic matter, particle size distribution, and $\log_{10}$Ksat
- There were a total of 323 unique input variable (Table 1,2)

Table 1. Soil series from the Southeastern US in the S-124 project.

Authors	Soils
Quisenberry, V., D. K. Cassel, J.H. Dane, and J.C. Parker	Norfolk, Dothan, Goldsboro, and Wagram
Römkens, M. J. M., H.M. Selim, H.D. Scott, R.E. Phillips, and F.D. Whisler	Captina, Gigger, Grenada, Loring, Olivier, and Sharkey
Nofziger, D., J.R. Williams, A.G. Hornsby, and A.L. Wood	Bethany, Konawa, and Tipton
Römkens, M.J.M., J.M. Selim, R.E. Phillips, and F.D. Whisler	Vicksburg, Memphis, and Maury
Bruce, R., J. Dane, V. Quisenberry, N. Powell, and A. Thomas	Cecil
R. Luxmoore	Fullerton and Sequoia
Dane, J., D.K. Cassel, J.M. Davidson, W.L. Pollans, and V.L. Quisenberry	Troup and Lakeland

Table 2. Summary of variables.

Group	Variable Name	Number of Members	Description
1-HOR	HOR	231(0) ; 92(1)	0 for subsoil ; 1 for topsoil
2-PED	PS1	70	1-2 mm ped size class
	PS2	30	2-5 mm ped size class
	PS3	15	5-10 mm ped size class
	PS4	116	10-20 mm ped size class
	PS5	36	20-50 mm ped size class
	PS6	26	50-100 mm ped size class
	PS7	11	>100 mm ped size class
	PS8	19	ped size not determined
3-CRK	BOTH	198	horizontal and vertical cracks
	TRANS	1	Horizontal cracks
	VERT	54	Vertical cracks
	MASSIVE	10	Massive
	SINGLE	45	Single
	NONE	15	Structureless
4-TXT	S	48	sand textural class (USDA)
	LS	31	Loamy sand textural class (USDA)
	SL	32	Sandy loam textural class (USDA)
	L	6	Loam textural class (USDA)
	ZL	78	Silt loam textural class (USDA)
	Z	2	Silt textural class (USDA)
	SCL	34	Sandy clay loam textural class (USDA)
	CL	30	Clay loam textural class (USDA)
	ZCL	27	Silty clay loam textural class (USDA)
	ZC	0	Silty clay textural class (USDA)
	SC	1	Sandy clay textural class (USDA)
	C	34	Clay textural class (USDA)
5-PSD	CLAY	323	Clay content
	SILT	323	Silt content
	SAND	323	Sand content
6-BDO	D(b)	323	Bulk density
	OM	N/A	Organic matter
Output	LOGKS	323	Transformed $K_{(sat)}$ $\log_{10}$ (cm d^{-1})

Results and Discussion

- Optimization parameters: five terminal nodes; 82% of samples for test dataset vs. 18% for validation dataset (Lilly et al., 2008); 63 unique variable inputs
- Table 3 shows all of the decision tree models listed in order of increasing RMSR
- Both trees had textural class sand as being the most important initial splitting variable and the highest Ksat (Figures 1,2)
- Figures 1 and 2 show larger ped sizes have smaller Ksat prediction
- Figure 2 shows that crack orientation plays no role in predicting Ksat even though it was one of the inputs

Table 3. Decision-tree models and significant variables.

Input Rank	--------Input Variables					--------	RMSR Mean $\log_{10}$(cm/day)	SD
1	PED			TXT	BDO		0.3695	0.0335
2	PED			TXT	BDO	HOR	0.3746	0.0339
3	PED	CRK		TXT	BDO		0.3775	0.0350
4		CRK		TXT	BDO	HOR	0.3806	0.0362
5	PED	CRK		TXT	BDO	HOR	0.3826	0.0345
6				TXT	BDO		0.3836	0.0352
7				TXT	BDO	HOR	0.3848	0.0355
8		CRK		TXT		HOR	0.3853	0.0396
9		CRK		TXT	BDO	HOR	0.3859	0.0357
10		CRK				HOR	0.3864	0.0354
11	PED			TXT		HOR	0.3882	0.0393
12	PED		PSD		BDO		0.3894	0.0380
13	PED		PSD		BDO	HOR	0.3894	0.0380
14	PED		PSD	TXT	BDO		0.3897	0.0379
15	PED		PSD	TXT	BDO	HOR	0.3898	0.0379
16	PED	CRK	PSD		BDO		0.3901	0.0379
17	PED	CRK	PSD		BDO	HOR	0.3901	0.0380
18	PED						0.3904	0.0378
19	PED	CRK	PSD	TXT	BDO	HOR	0.3905	0.0379
20			PSD	TXT		HOR	0.3910	0.0293
21	PED		PSD	TXT			0.3914	0.0319
22				TXT		HOR	0.3915	0.0400
23	PED	CRK			BDO	HOR	0.3917	0.0314
24	PED	CRK					0.3930	0.0324
25	PED	CRK				HOR	0.3955	0.0350
26	PED	CRK				HOR	0.3963	0.0377
27		CRK			BDO		0.3965	0.0299
28		CRK	PSD		BDO		0.3980	0.0381
29		CRK					0.3982	0.0353
30		CRK	PSD	TXT	BDO	HOR	0.3982	0.0389
31		CRK	PSD		BDO	HOR	0.3982	0.0389
32		CRK	PSD	TXT	BDO		0.3985	0.0382
33				TXT	BDO	HOR	0.3987	0.0384
34			PSD		BDO		0.3987	0.0384
35			PSD	TXT	BDO	HOR	0.3990	0.0391
36			PSD		BDO	HOR	0.3990	0.0391
37	PED			TXT			0.4014	0.0392
38			PSD	TXT			0.4047	0.0463
39			PSD				0.4047	0.0463
40			PSD			HOR	0.4054	0.0462
41		CRK	PSD	TXT			0.4054	0.0462
42		CRK	PSD				0.4062	0.0451
43		CRK	PSD	TXT		HOR	0.4062	0.0452
44	PED	CRK		TXT			0.4065	0.0448
45				TXT			0.4065	0.0448
46		CRK		TXT			0.4079	0.0343
47	PED	CRK	PSD	TXT		HOR	0.4084	0.0404
48	PED	CRK	PSD	TXT			0.4096	0.0383
49	PED	CRK	PSD				0.4123	0.0474
50	PED		PSD			HOR	0.4123	0.0474
51	PED		PSD				0.4124	0.0477
52	PED		PSD	TXT		HOR	0.4124	0.0477
53	PED		PSD	TXT			0.4124	0.0483
54	PED				BDO	HOR	0.4125	0.0485
55	PED		PSD	TXT		HOR	0.4129	0.0486
56	PED		PSD	TXT			0.4129	0.0489
57	PED				BDO	HOR	0.4144	0.0350
58	PED					HOR	0.4179	0.0404
59	PED				BDO		0.4211	0.0341
60	PED						0.4317	0.0388
61					BDO	HOR	0.4389	0.0330
62					BDO		0.4444	0.0306
63						HOR	0.4835	0.0359

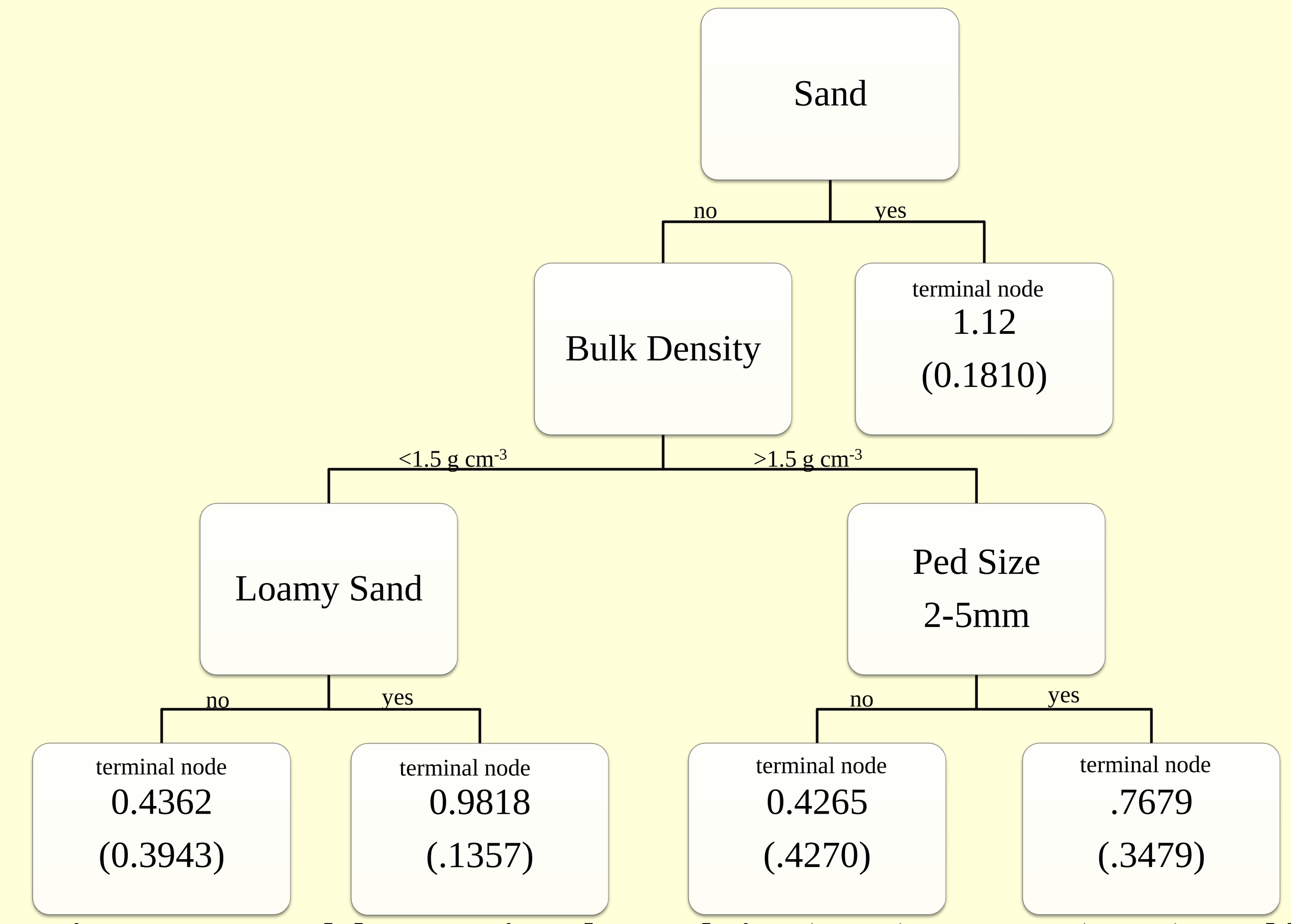

Fig. 1. Best model run, using the ped size (PED), texture (TXT), and bulk density (BDO) inputs.

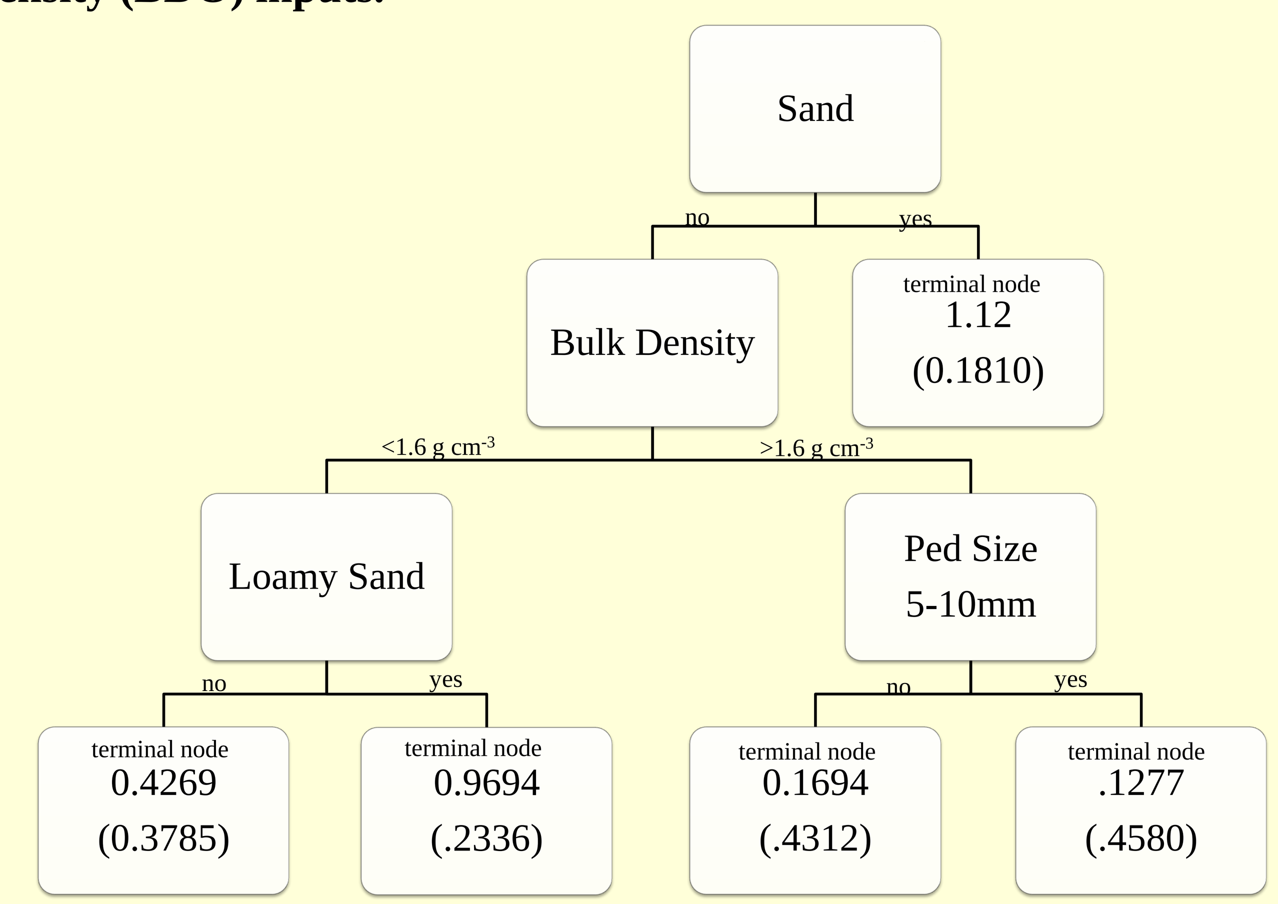

Fig. 2. Fifth ranked model, using the ped size (PED), crack orientation (CRK), texture (TXT) , bulk density (BDO), and horizon (HOR) inputs.

Conclusions

- The structure variable ped size was most important in both decision tree models
- Ksat predictions for sand in decision tree models are low
- HOR from the 5th ranked model has little effect on the predictive capability of the model

References

Lilly, A., A. Nemes, W.J. Rawls, and Ya. A. Pachepsky. 2008. Probabilistic approach to the identification of input variable to estimate hydraulic conductivity. Soil Sci. Soc. Am. J. 72:16-24.