Table 2. Performance of *k*-fold cross-validation between benchmark (IPCC Tier 1) and knife set (KS) approach

		IPCC Tier 1		Knife Set (KS) approach	
		Mean	SD	Mean[2,3]	SD
mean absolute percentage error (MAPE)	AGC[1] (%)	39.75%	**23.90%**	**23.29%*** **	24.74%
	BGC (%)	122.12%	140.90%	**41.84%** **	**25.15%**
	TC (%)	34.32%	23.70%	**26.22%** •	**15.61%**
	ANPP (%)	172.38%	250.90%	**48.92%** **	**38.53%**
	BNPP (%)	505.34%	635.06%	**59.30%***	**65.84%**
	TNPP (%)	328.44%	390.94%	**52.35%***	**51.45%**
root mean square error (RMSE)	AGC (Mg C ha^{-1})	31.84	32.45	**14.37*****	**19.75**
	BGC (Mg C ha^{-1})	18.67	22.14	**12.96***	**13.50**
	TC (Mg C ha^{-1})	30.86	33.53	**20.03** **	**24.21**
	ANPP (Mg C ha^{-1})	6.64	10.91	**4.48**	**6.88**
	BNPP (Mg C ha^{-1})	4.16	5.7	**3.12**	**4.22**
	TNPP (Mg C ha^{-1})	13.41	16.97	**8.15**	**10.36**
mean absolute error (MAE)	AGC (Mg C ha^{-1})	27.92	18.83	**11.50*****	**13.61**
	BGC (Mg C ha^{-1})	13.9	12.63	**11.03*****	**6.89**
	TC (Mg C ha^{-1})	25.08	18.21	**16.87***	**10.93**
	ANPP (Mg C ha^{-1})	4.11	5.29	**3.16**	**3.22**
	BNPP (Mg C ha^{-1})	3.05	2.97	**2.25**	**2.27**
	TNPP (Mg C ha^{-1})	9.63	9.79	**5.79**	**6.02**
mean error (ME)	AGC (Mg C ha^{-1})	-19.39	27.62	**-2.40*****	**17.70**
	BGC (Mg C ha^{-1})	12.02	14.47	**-0.45*****	**13.12**
	TC (Mg C ha^{-1})	-11.05	29.18	**-3.85**	**19.91**
	ANPP (Mg C ha^{-1})	3.28	5.86	**0.38***	**4.53**
	BNPP (Mg C ha^{-1})	3.02	**3**	**-0.02***	3.27
	TNPP (Mg C ha^{-1})	9.63	9.79	**3.46**	**7.74**

[1] Aboveground carbon (AGC) (n=77), belowground carbon (BGC) (n=39), total carbon (TC) (n=40), aboveground net primary production (ANPP) (n=33), belowground net primary production (BNPP) (n=11), and total net primary production (TNPP) (n=11).

[2] Means with different significant codes (*** $p \leqq 0.001$, ** $p \leqq 0.01$, * $p \leqq 0.05$, • $p \leqq 0.1$) indicate that performance of the KS approach is significantly lower than benchmark by one-way ANOVA or Welch's Heteroscedastic F Test depending on that observation is homogeneity or heterogeneity.

[3] Bold black colour presents lower values (higher performance).