

**Table S1.** Statistical significance of the difference in OA between the ten models compared using McNemar's test (significant differences with p-value < 0.05 are indicated with \*). McNemar test statistic values are on the left side of the diagonal, and p-values on the right side.

	PB_S2	PB_S2S1	PB_PL	PB_S2S1PL	OB_S2	OB_S2S1	OB_S2S1T	OB_PL	OB_S2S1PL	OB_PLS2S1
PB_S2		0.003	0.014	0.005	0.005	0.001	0.000	0.006	0.000	0.000
PB_S2S1	8.889 *		0.000	1.000	1.000	0.401	0.033	0.804	0.002	0.000
PB_PL	6.011 *	19.959 *		0.000	0.000	0.000	0.000	0.000	0.000	0.000
PB_S2S1PL	7.875 *	0	25.312*		1.000	0.48	0.055	0.896	0.002	0.000
OB_S2	7.875 *	0	22.5 *	0		0.327	0.026	0.892	0.004	0.000
OB_S2S1	11.045 *	0.706	29.557 *	0.5	0.962		0.118	0.651	0.041	0.002
OB_S2S1T	21.121 *	4.558 *	37.837 *	3.674	4.971 *	2.450		0.105	0.337	0.040
OB_PL	7.557 *	0.062	31.557 *	0.017	0.019	0.205	2.630		0.004	0.000
OB_S2S1PL	28 *	9.587 *	48.944 *	9.756 *	8.163 *	4.17 *	0.923	8.308 *		0.23
OB_PLS2S1	35.309 *	15.373 *	59.211 *	16.568*	16.568 *	9.587 *	4.225 *	14.881 *	1.44	