Model Selection in Historical Biogeography Reveals that Founder-event Speciation is a Crucial Process in Island Clades

Nicholas J. Matzke[1, 2]

[1] Department of Integrative Biology, University of California, Berkeley

[2] National Institute for Mathematical and Biological Synthesis and Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN. Email: *matzke@nimbios.org*.

Supplemental Text: Supplemental Methods, Results, and Discussion

Supplemental Methods

Tree digitization methods

Where possible, original tree files were gathered from supplemental material or other online sources, but usually the only available source for the tree was the published graphic. As has been noted (Boettiger and Lang 2012), the digital archiving of published, dated, phylogenies is still not standard practice, and the limited tree-digitization software available is typically platform-specific or otherwise difficult to use, driving some researchers to use calipers on the published trees in printed-out journal article pages in order obtain trees with branchlengths (Boettiger and Lang 2012). Here, a relatively workable alternative approach to digitizing trees was discovered. GraphClick, a cheap and user-friendly manual digitization program (Arizona Software, $8, http://www.arizona-software.ch/graphclick/) was used to obtain the *x* and *y* coordinates of the tip nodes, the internal nodes, and the "corners" (the branch bottoms below nodes on a

typical phylogeny with square corners). Then, an R script, TreeRogue (Matzke 2013) was used to assemble the coordinates and tip labels into a standard Newick file. An experienced user can use this system to digitize a tree to Newick format in a matter of minutes; some of the trees used in this study were digitized by undergraduate research assistants, although this requires above-average computer skills, training in the reading and interpretation of phylogenies, and strong attention to detail, as missing even one node during digitization will cause the Newick conversion operation to fail. Accuracy of digitization is approximately 1 millimeter on a printed page or computer screen, which translates to less than 1% error in the digitized tree compared to the original, certainly well below the statistical uncertainty inherent in any dating analysis.

Empirical datasets

The clades and source studies were (1) Hawaiian *Psychotria* (Nepokroeff et al. 2003; Ree and Smith 2008), the example dataset for the original Python LAGRANGE (Ree and Smith 2008); (2) Hawaiian *Drosophila*, a widely-cited example of founder-event speciation (Carson 1974; Hampton and Kaneshiro 1976; Templeton 1980; DeSalle and Templeton 1988) as sampled by Kambysellis et al. (1995); (3) *Scaptomyza* (Lapoint et al. 2013); (4) Hawaiian honeycreepers (Lerner et al. 2011); (5) Hawaiian damselflies, *Megalagrion* (Jordan 2003; Jordan et al. 2003); (6) Hawaiian leafhoppers, *Nesophrosyne* (Bennett and O'Grady 2011; Bennett and O'Grady 2013); (7) *Orsonwelles* (Araneae; Hormiga 2002) a spider genus endemic to Hawaii (Hormiga et al. 2003); (8) Hawaiian *Plantago* (Dunbar-Co

2008); (9) the Hawaiian silversword alliance (Asteraceae) with the tree and ranges assembled from two papers (Baldwin and Sanderson 1998; Gillespie and Baldwin 2010); (10) Caribbean and Central/South American *Anolis* (Köhler and Vesely 2010; Nicholson et al. 2012); in this case the dated tree was re-estimated in BEAST using methods described in Supplemental Text with data provided by Nicholson (personal communication); (11) *Microlophus*, the "lava lizards" of the Galapagos (Benavides et al. 2009); (12) Pacific *Cyrtandra* (Gesneriaceae), previously examined by Clark et al. (Clark et al. 2008; Clark et al. 2009) in their comparative study of biogeography methods; (13) Southeast Asian members of *Rhododendron* subgenus *Vireya* (Brown et al. 2006; Webb and Ree 2012); this served as the test clade for SHIBA (Webb and Ree 2012) and BayArea (Landis et al. 2013). Like Caribbean *Anolis*, but unlike the other clades considered in this study, *R.* subgen. *Vireya* is not primarily found on oceanic islands; it has representatives spread from the Himalayas to Australia (Webb and Ree 2012). However, Malesia is its center of diversity, and it is interesting to test whether or not a clade distributed on continental islands will support the DEC or DEC+J models in a fashion similar to oceanic island clades.

Estimation of the Caribbean Anolis *tree*

The dated tree of *Anolis* (Nicholson et al. 2012) was regenerated from scratch by taking the input sequence alignment (obtained courtesy of K. Nicholson, personal communication) and conducting a BEAST dating analysis (Drummond and Rambaut 2007; Drummond et al. 2012) using the settings and calibration points described by Nicholson (2012), and taking the resulting well-resolved maximum clade credibility

tree (MCC) for use in biogeographical analysis. It should be noted that the dates obtained by Nicholson et al. (2012) are quite old (the root of the genus tree is 87 Ma), as a result of the input calibrations, which consist of two fossils of age 28 mya and 17-23 mya, each of which the authors place very high in the tree as sister to extant species or small clades. Given the recent documentation of extensive morphological convergence in *Anolis* (Mahler et al. 2013) these calibrations might be questioned, especially as Nicholson et al. note that many workers in the group favor younger dates (Nicholson et al. 2012). Nicholson *et al*. premise their discussion of biogeography on ancient dates and vicariance interacting with the complex geological history of the Caribbean. However, their biogeographic inference method is parsimony character mapping combined with narrative interpretation; there is therefore no explicit dispersal matrix or other constraint that must be taken into account in the present likelihood analysis (such an analysis would a fascinating, but very large, project). Therefore, only an unconstrained, non-time-stratified analysis was run in the present study.

R. *subgen.* Vireya *dating*

Due to the lack of rigorous dating information for the clade, Webb & Ree (2012) explore two different scalings of the age of the clade, 11 and 55 my. As a time-stratified analysis is not being reproduced in this study, this scaling issue is unimportant for model selection, so only the 11 my scaling was used.

Bayesian inference under DEC and DEC+J: Extended Methods

ML estimation only provides point estimates of parameters, and frequentist model tests such as the LRT require certain theoretical assumptions that may not always be met in real datasets. As a check on the likelihood-based model selection procedures used for most of this study, a Bayesian version of inference and model choice under DEC and DEC+J inference was implemented on the test *Psychotria* dataset, unconstrained (M0) geography model.

The R package LaplacesDemon (Statisticat 2013b, available at http://www.bayesian-inference.com/software) implements a wide variety of Bayesian techniques. Given only user-specified priors on free parameters, and a user-specified likelihood function, MCMC chains can be constructed with minimal overhead in LaplacesDemon, which also supplies sophisticated diagnostic functions and plots. LaplacesDemon implements more than two dozen MCMC algorithms, two of which were used here. First, standard random-walk Metropolis MCMC (Link and Barker 2009; Statisticat 2013a) was employed. Uniform priors were applied to the *d*, *e*, and *j* parameters, with the bounds set to match the limits used in ML search (0-5 for *d* and *e*, and 0-3 for *j*). The same BioGeoBEARS function that was used to calculate the data likelihood in ML estimation was used for the Bayesian analysis. The DEC and DEC+J models were run separately, each for 10,000 generations. A 2- or 3-parameter problem is a simple one compared to many MCMC analyses, and inspection of trace and autocorrelation plots indicated that this run length was more than sufficient to accomplish burnin (which occurred within 200 generations), establish stationarity, and achieve sufficient sampling to characterize the posterior.

The posterior samples were used to plot the posterior distribution of each parameter under each model. To approximate the marginal log-likelihood of each model for Bayesian model choice, the nonparametric self-normalized importance sampling algorithm of Escoto (2011) was used (LaplacesDemon LML function, NSIS option). The Bayes Factor was calculated from the ratio of the approximate marginal likelihoods.

The second MCMC analysis used reversible-jump MCMC (Green 1995) with both DEC and DEC+J sampled in the same MCMC search. As is often found in RJMCMC analyses, due to the strong likelihood advantage of one model (in this case, DEC+J), the prior probability of the two models has to be extremely heavily biased in favor of the poorer model (DEC) in order to achieve any sampling at all of the weaker model (Link and Barker 2009). A prior that produced adequate sampling of both models was thus sought by trial and error, resulting in a prior probability of 0.000001 placed on the DEC+J model. Not coincidently, this prior almost exactly balances the log-likelihood advantage of DEC+J (~14 log-likelihood units) on the *Psychotria* dataset. Priors even slightly different resulted in runs that exclusively or almost exclusively sampled only one model. The functioning RJMCMC analysis was run for 50,000 generations and was assessed as described above. Here, the Bayes factor was calculated as the ratio of the posterior probability of DEC+J (which is simply the frequency at which the DEC+J model was sampled in the post-burnin posterior distribution) to the prior probability of the DEC+J model.

Supplemental Results

Bayesian inference on Psychotria *M0 dataset: Results*

Trace plots and histograms for the parameters and the model log-posterior probabilities are shown in Supplemental Figures 1.1-1.6. The results of random-walk Metropolis MCMC are shown in Supplemental Figures 1.1-1.4. Inference of *d* appears to be unproblematic for both DEC and DEC+J, with the ML estimate of *d* appearing at the peak of the posterior distribution in both cases. However, for the *e* parameter, the ML estimate (0.028) is well outside the 95% highest posterior distribution (HPD) for the DEC model, a result further highlighting the weaknesses of DEC inference in estimating *e*. Even when the ML estimate of *e* is nonzero, the likelihood surface is extremely broad, and the likelihood advantage of a positive *e* value over *e*=0 is tiny. Inference of *j* appears unproblematic, with the ML estimate of *j* (0.11) within the 95% HPD, and with *j*=0 outside of the 95% HPD.

The estimates of the log-marginal likelihood for DEC and DEC+J were -44.6 and -36.4, respectively. This yields a Bayes Factor of 3595 in favor of DEC+J. On the Jeffreys' scale for the interpretation of Bayes Factors (Jeffreys 1961), anything above 100:1 is considered decisive support for a model. Under reversible-jump MCMC, prior probability of DEC+J was set to 0.000001, and the posterior probability was 0.1524, resulting in a Bayes Factor of 152422. This Bayes Factor is similar in magnitude to the ratio of model weights calculated from AIC, and indicates that the likelihood advantage of DEC+J is dominating the result. The Bayes Factor calculated from the posterior distribution random-walk Metropolis MCMC appears to be more conservative, taking into account the uncertainty in parameter inference and resulting uncertainty in model choice.

Supplemental Text References

Baldwin BG, Sanderson MJ. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proceedings of the National Academy of Sciences, 95:9402-9406.

Benavides E, Baum R, Snell HM, Snell HL, Sites JW, Jr. 2009. Island Biogeography of Galápagos Lava Lizards (Tropiduridae: Microlophus): Species Diversity and Colonization of the Archipelago. Evolution, 63:1606-1626.

Bennett GM, O'Grady PM. 2013. Historical biogeography and ecological opportunity in the adaptive radiation of native Hawaiian leafhoppers (Cicadellidae: *Nesophrosyne*). Journal of Biogeography, 40:1512-1523.

Bennett GM, O'Grady PM. 2011. Review of the native Hawaiian leafhopper genus *Nesophrosyne* (Hemiptera: Cicadellidae: Deltocephalinae) with description of eight new species associated with *Broussaisia arguta* (Hydrangeaceae). Zootaxa, 2805:1-25.

Boettiger C, Lang DT. 2012. Treebase: an R package for discovery, access and manipulation of online phylogenies. Methods in Ecology and Evolution, 3:1060-1066.

Brown GK, Nelson G, Ladiges PY. 2006. Historical biogeography of Rhododendron section *Vireya* and the Malesian Archipelago. Journal of Biogeography, 33:1929-1944.

Carson HL. 1974. Patterns of speciation in Hawaiian *Drosophila* inferred from ancient chromosomal polymorphism. In: White MJD editor. Genetic Mechanisms of Speciation in Insects, Springer Netherlands, p. 81-93.

Clark JR, Ree RH, Alfaro ME, King MG, Wagner WL, Roalson EH. 2008. A comparative study in ancestral range reconstruction methods: retracing the uncertain histories of insular lineages. Systematic Biology, 57:693-707.

Clark JR, Wagner WL, Roalson EH. 2009. Patterns of diversification and ancestral range reconstruction in the southeast Asian–Pacific angiosperm lineage *Cyrtandra* (Gesneriaceae). Molecular Phylogenetics and Evolution, 53:982-994.

DeSalle R, Templeton AR. 1988. Founder Effects and the Rate of Mitochondrial DNA Evolution in Hawaiian *Drosophila*. Evolution, 42:1076-1084.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology, 7:214.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution, 29:1969-1973.

Dunbar-Co S (2008) Island Evolution: Phylogeny, Adaptive Radiation, and Biogeography of *Plantago* (Plantaginaceae) in the Hawaiian Islands. Ph.D., University of Hawai'i.

Escoto B. 2011. Marginal Likelihood Computation via Arrogance Sampling. arXiv.org:1-12.

Gillespie RG, Baldwin BG. 2010. Island biogeography of remote archipelagos: Interplay between ecological and evolutionary processes. In: Losos JB, Ricklefs RE, MacArthur RH editors. The theory of island biogeography revisited. Princeton, Princeton University Press, p. 358-387.

Green PJ. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika, 82:711-732.

Hampton LC, Kaneshiro KY. 1976. *Drosophila* of Hawaii: Systematics and Ecological Genetics. Annual Review of Ecology and Systematics, 7:311-345.

Hormiga G. 2002. *Orsonwelles*, a new genus of giant linyphiid spiders (Araneae) from the Hawaiian Islands. Invertebrate Systematics, 16:369-448.

Hormiga G, Arnedo M, Gillespie RG. 2003. Speciation on a Conveyor Belt: Sequential Colonization of the Hawaiian Islands by *Orsonwelles* Spiders (Araneae, Linyphiidae). Systematic Biology, 52:70-88.

Jeffreys H. 1961. The theory of probability. Oxford, Oxford University Press.

Jordan S, Simon C, Polhemus D. 2003. Molecular Systematics and Adaptive Radiation of Hawaii's Endemic Damselfly Genus *Megalagrion* (Odonata: Coenagrionidae). Systematic Biology, 52:89-109.

Kambysellis MP, Ho K-F, Craddock EM, Piano F, Parisi M, Cohen J. 1995. Pattern of ecological shifts in the diversification of Hawaiian *Drosophila* inferred from a molecular phylogeny. Current Biology, 5:1129-1139.

Köhler G, Vesely M. 2010. A Revision of the *Anolis sericeus* Complex with the Resurrection of *A. wellbornae* and the Description of a New Species (Squamata: Polychrotidae). Herpetologica, 66:207-228.

Landis M, Matzke NJ, Moore BR, Huelsenbeck JP. 2013. Bayesian Analysis of Biogeography when the Number of Areas is Large. Systematic Biology.

Lapoint RT, O'Grady PM, Whiteman NK. 2013. Diversification and dispersal of the Hawaiian Drosophilidae: The evolution of *Scaptomyza*. Molecular Phylogenetics and Evolution, 69:95-108.

Lerner Heather RL, Meyer M, James Helen F, Hofreiter M, Fleischer Robert C. 2011. Multilocus Resolution of Phylogeny and Timescale in the Extant Adaptive Radiation of Hawaiian Honeycreepers. Current Biology, 21:1838-1844.

Link WA, Barker RJ. 2009. Bayesian Inference: with ecological applications. Academic Press.

Mahler DL, Ingram T, Revell LJ, Losos JB. 2013. Exceptional Convergence on the Macroevolutionary Landscape in Island Lizard Radiations. Science, 341:292-295.

Matzke NJ (2013) TreeRogue: R code for digitizing trees. https://stat.ethz.ch/pipermail/r-sig-phylo/2010-October/000816.html

Nepokroeff M, Sytsma KJ, Wagner WL, Zimmer EA. 2003. Reconstructing Ancestral Patterns of Colonization and Dispersal in the Hawaiian Understory Tree Genus *Psychotria* (Rubiaceae):A Comparison of Parsimony and Likelihood Approaches. Systematic Biology, 52:820-838.

Nicholson KE, Crother BI, Guyer C, Savage JM. 2012. It is time for a new classification of anoles (Squamata: Dactyloidae). Zootaxa, 3477:1-108.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology, 57:4-14.

Statisticat L (2013a) LaplacesDemon Tutorial. http://www.bayesian-inference.com/softwaredownload

Statisticat L (2013b) LaplacesDemon: Complete Environment for Bayesian Inference. http://www.bayesian-inference.com/softwaredownload

Templeton AR. 1980. The theory of speciation via the founder principle. Genetics, 94:1011-1038.

Webb CO, Ree RH. 2012. Historical biogeography inference in Malesia. In: Gower D, Johnson K, Richardson J, Rosen B, Ruber L, Williams S editors. Biotic evolution and environmental change in Southeast Asia. Cambridge, Cambridge University Press, p. 191-215.

Wood HM, Matzke NJ, Gillespie RG, Griswold CE. 2013. Treating Fossils as Terminal Taxa in Divergence Time Estimation Reveals Ancient Vicariance Patterns in the Palpimanoid Spiders. Systematic Biology, 62:264-284.

Supplementary Table 1. Model selection statistics comparing DEC and DEC+J on 13 island clades, for 53 constraints scenarios. The interpretation of the *p*-values is as follows: *, p<0.05; **, *p*<0.01; ***, p<0.001. *Abbreviations:* strat., time-stratified; disp., manual dispersal probability multiplier matrix; w. Z, analysis was run with "Z", an old, ancestral area outside of the extant Hawaiian high islands; nz, zeros in the manual dispersal probability multiplier matrix have been replaced with a small nonzero value.

Clade	region	tree age	# tips	Constraints model ID	Description	DEC+J LnL	DEC LnL	LRT pval	sig.	AIC weight ratio	AICc weight ratio
Psychotria	Hawaii	5.2	19	M0	unconstrained	-20.9	-34.5	1.8E-07	***	294895	191952
Psychotria	Hawaii	5.2	19	M1	2 areas max	-21.0	-35.2	9.4E-08	***	566941	371387
Psychotria	Hawaii	5.2	19	M2	2 areas + east disp. only	-16.7	-31.9	3.4E-08	***	1.5E+06	1.0E+06
Psychotria	Hawaii	5.2	19	M2b	4 areas max, east-only disp.	-16.7	-31.4	5.7E-08	***	924075	606221
Psychotria	Hawaii	5.2	19	M3	strat. disp.	-28.3	-39.8	1.6E-06	***	36882	23981
Psychotria	Hawaii	5.2	19	M3b	strat. disp. & areas	-32.8	-41.7	2.4E-05	***	2771	1817
Hawaiian *Drosophila*	Hawaii	20	42	M0	unconstrained	-59.5	-97.1	4.6E-18	***	7.3E+15	6.2E+15
Hawaiian *Drosophila*	Hawaii	20	42	M3areas	strat. disp. & areas	-224.9	-226.2	0.11	ns	1.3	1.1
Hawaiian *Drosophila*	Hawaii	20	42	M3areas	strat. disp. & areas, w. Z	-547.6	-551.2	0.0068	**	14.3	11.5
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M0	unconstrained	-61.9	-103.0	1.2E-19	***	2.6E+17	2.3E+17
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M1	2 areas max	-62.2	-102.3	3.4E-19	***	9.4E+16	8.5E+16
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M1a	2 areas max; adjacency	-62.4	-121.7	1.4E-27	***	2.0E+25	1.7E+25
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M2	2 areas max; east disp. only	-116.4	-133.1	7.6E-09	***	6.5E+06	5.7E+06
Hawaiian *Scaptomyza*	Hawaii	12.4	49	M3b_stratified_w_areas	strat. disp. & areas	-218.9	-224.0	0.0013	**	64.8	52.7
Hawaiian *Scaptomyza*5	Hawaii	12.4	49	M0	unconstrained, w. Z	-68.8	-108.9	3.4E-19	***	9.5E+16	8.3E+16
Hawaiian *Scaptomyza*5	Hawaii	12.4	49	M1	2 areas max, w. Z	-69.2	-108.1	1.2E-18	***	2.7E+16	2.4E+16
Hawaiian *Scaptomyza*5	Hawaii	12.4	49	M1a	2 areas max; adjacency, w. Z	-69.5	-128.5	1.7E-27	***	1.5E+25	1.3E+25
Hawaiian *Scaptomyza*5	Hawaii	12.4	49	M2	2 areas max; east disp. only, w. Z	-106.9	-119.6	4.9E-07	***	1.1E+05	1.1E+05
Hawaiian *Scaptomyza*5	Hawaii	12.4	49	M3b_stratified_w_areas	strat. disp. & areas, w. Z	-273.2	-273.5	0.41	ns	0.5	0.4
Hawaiian *Scaptomyza*5	Hawaii	12.4	49	M3c_stratified_w_areas	strat. disp. & areas, w. Z, low prob.	-505.6	-510.4	0.002	**	44.4	39.0
Hawaiian Honeycreepers	Hawaii	5.78	19	M0	unconstrained	-43.1	-55.2	8.6E-07	***	67028	43695
Hawaiian Honeycreepers	Hawaii	5.78	19	M3_strat	strat. disp.	-59.9	-64.5	0.0026	**	34.8	22.8
Hawaiian Honeycreepers	Hawaii	5.78	19	M3a_strat	strat. disp. + areas	-66.9	-78.7	1.2E-06	***	49078	32048
Megalagrion	Hawaii	9.6	34	M0	unconstrained	-43.4	-69.5	5.1E-13	***	7.8E+10	6.3E+10
Megalagrion	Hawaii	9.6	34	M1_2areas	2 areas max	-43.6	-72.0	4.9E-14	***	7.8E+11	6.3E+11
Megalagrion	Hawaii	9.6	34	M1a_2areas	2 areas max; adjacency	-44.0	-77.7	2.3E-16	***	1.5E+14	1.2E+14
Megalagrion	Hawaii	9.6	34	M2_2areasEastOnly	2 areas max; east disp. only	-46.4	-70.4	4.2E-12	***	9.8E+09	8.0E+09
Megalagrion	Hawaii	9.6	34	M3b_stratified_w_areas	strat. disp. & areas	-79.9	-90.0	7.4E-06	***	8518	6930
Megalagrion	Hawaii	9.6	34	M0	unconstrained	-43.4	-69.5	5.1E-13	***	7.8E+10	6.3E+10
Megalagrion	Hawaii	9.6	34	M1_2areas	2 areas max	-43.6	-72.0	4.9E-14	***	7.8E+11	6.3E+11
Megalagrion	Hawaii	9.6	34	M1a_2areas	2 areas max; adjacency	-44.0	-77.7	2.3E-16	***	1.5E+14	1.2E+14
Megalagrion	Hawaii	9.6	34	M2_2areasEastOnly	2 areas max; east disp. only	-46.4	-70.4	4.2E-12	***	9.8E+09	8.0E+09
Megalagrion	Hawaii	9.6	34	M3b_stratified_w_areas	strat. disp. & areas	-79.9	-90.0	7.4E-06	***	8518	6930
Megalagrion	Hawaii	9.6	34	M3c_stratified_w_areas	strat. disp. & areas, w. Z	-136.0	-145.3	1.5E-05	***	4303	3273
Nesophrosyne	Hawaii	3.2	198	M1 (UM-2)	unconstrained	-255.7	-371.6	2.4E-52	***	8.0E+49	7.7E+49
Nesophrosyne	Hawaii	3.2	198	M1 (UM-2)	2 areas max	-255.8	-376.9	1.3E-54	***	1.5E+52	1.4E+52
Nesophrosyne	Hawaii	3.2	198	M3_6max (TS-U)	unconstrained, strat. disp.	-360.5	-434.8	3.4E-34	***	7.0E+31	6.6E+31
Nesophrosyne	Hawaii	3.2	198	M3_2max (TS-2)	2 areas max; strat. disp. & areas	-497.4	-528.1	4.9E-15	***	7.5E+12	7.7E+12
Nesophrosyne	Hawaii	3.2	198	M3_6max (TS-U)	unconstrained, strat. disp. & areas	-500.2	-537.2	8.3E-18	***	4.1E+15	4.2E+15
Nesophrosyne	Hawaii	3.2	198	M3_6max (TS-Ua)	unconstrained, strat. disp. (nz) & areas	-506.1	-543.4	5.7E-18	***	5.9E+15	5.6E+15
Nesophrosyne	Hawaii	3.2	198	M3_2max (TS-2)	2 areas max; strat. disp. (nz) & areas	-499.8	-534.3	1.0E-16	***	3.5E+14	3.4E+14
Orsonwelles	Hawaii	4.35	12	M0	unconstrained	-13.1	-15.8	0.019	*	5.7	2.5
Plantago	Hawaii	5.1	16	M0	unconstrained	-15.8	-23.2	1.0E-04	***	625	366
Plantago	Hawaii	5.1	16	M3_strat	strat. disp.	-22.0	-25.3	0.01	*	9.8	5.8
Plantago	Hawaii	5.1	16	M3_strat	strat. disp. and areas	-41.5	-44.7	0.012	*	8.7	5.1
Silversword alliance	Hawaii	6.84	29	M0	unconstrained	-48.9	-51.3	0.027	*	4.2	3.3
Silversword alliance	Hawaii	6.84	29	M3_strat	strat. disp., with areas	-53.5	-57.9	0.0032	**	28.8	22.4
Anolis	Caribbean	98.9	190	M0	unconstrained	-242.0	-275.9	1.8E-16	***	1.9E+14	1.9E+14
Cyrtandra	Pacific	41.5	59	M0	unconstrained	-42.5	-47.3	0.0019	**	44.8	40.4
Cyrtandra	Pacific	41.5	59	M1	4 areas max	-42.5	-47.3	0.0019	**	44.8	40.4
Microlophus	Galapagos	3.75	9	M1	2 areas max	-20.5	-34.8	8.2E-08	***	6.5E+05	1.6E+05
Vireya	SE Asia	11	65	M1	2 areas, no other constraints	-214.1	-250.9	1.0E-17	***	3.3E+15	3.2E+15
Vireya	SE Asia	11	65	M1+cc	2 areas + constrained connectivity	-199.7	-219.4	3.6E-10	***	1.3E+08	1.2E+08

Supplemental Figure 1. Bayesian Analysis: parameter inference and MCMC diagnostic plots.

Supplemental Figure 1.1. Inference of parameters for the Psychotria M0 dataset, under the DEC model, using random-walk Metropolis MCMC.

Supplemental Figure 1.2. Trace and histogram of the log-posterior probabilities (LP) of the DEC model, using random-walk Metropolis MCMC.

Supplemental Figure 1.3. Inference of parameters for the Psychotria M0 dataset, under the DEC+J model, using random-walk Metropolis MCMC.

Supplemental Figure 1.4. Trace and histogram of the log-posterior probabilities (LP) of the DEC+J model, using random-walk Metropolis MCMC.

Supplemental Figure 1.5. Inference of parameters for the Psychotria M0 dataset, using reversible-jump MCMC to sample both the DEC and the DEC+J model.

Supplemental Figure 1.6. Trace and histogram of the log-posterior probabilities (LP) when using reversible-jump MCMC to sample both the DEC and the DEC+J model.

Supplemental Figure 2. The log-likelihood advantage of DEC+J inference over DEC, when the true model is DEC (top) or DEC+J (bottom).

Supplemental Figure 3. Histograms of the accuracies of ancestral state inference for each pair of simulation and inference under DEC and DEC+J. Accuracy equals the fraction of nodes for which the most probable inferred ancestral state matches the true, simulated ancestral state.

Supplemental Figure 4. Mean *S* for each node age, plotted against node age. *S* equals mean similarity to truth, i.e. 1 minus the absolute difference in the true probability of the true state (which is always 1) and the inferred probability of that state. The x-coordinates for the same node have been jittered to reduce overlap.

Supplemental Figure 5. Accuracy of parameter inference under DEC and DEC+J, when the history was simulated under each of these models. The true parameter used in the simulations is shown by the dashed line. These parameter values were the ML estimates in the original M0 analysis of *Psychotria* under DEC and DEC+J.

Supplemental Figures 6.1-6.6. Parameter values, ancestral range accuracy, and model choice under DEC and DEC+J inference, for a variety of *d*, *e*, and *j* values, under each of the size macroevolutionary processes generating the tree. These simulations started with a lineage with a geographic range of a single area. Parameter combinations and results are shown in 23 columns per page. A subset of these simulation results are shown in Figures 3 and 4, main text.

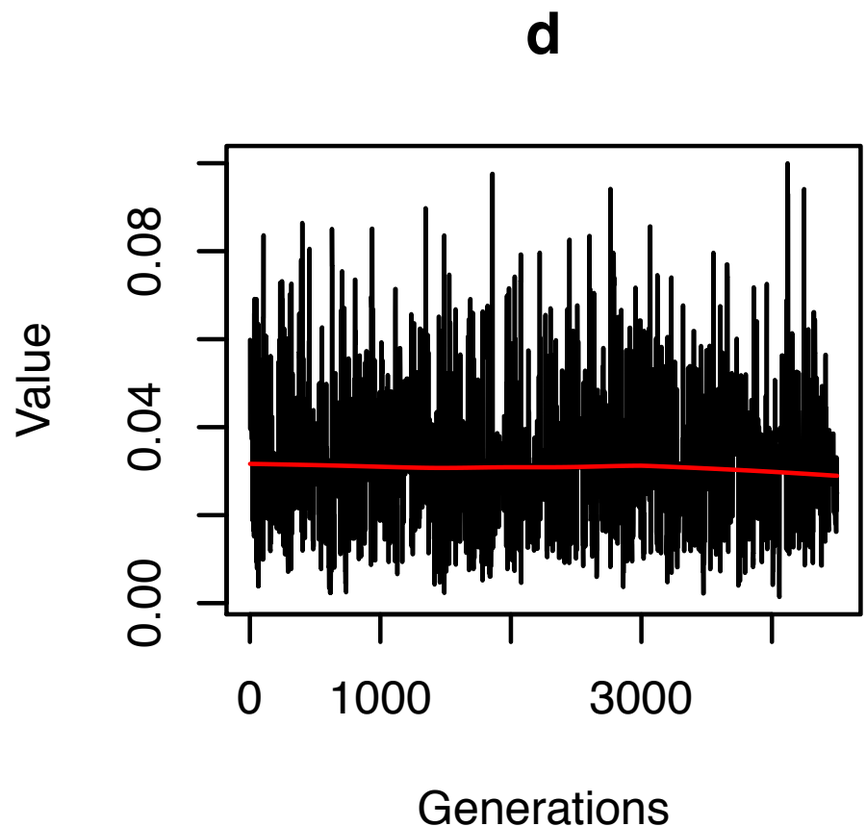
Supplemental Figures 7.1-7.6. Results plotted as in Supplemental Figure 6, for simulations that began with a lineage with a geographic range of all four areas.

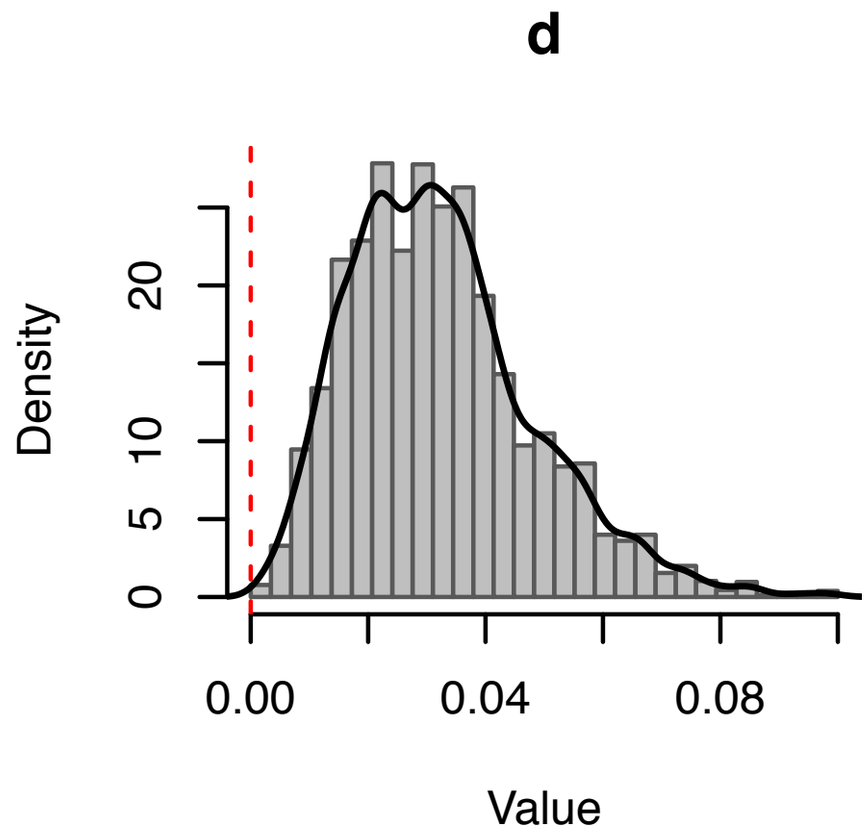
Supplemental Figures 8.1-8.12. Summary statistics for the trees and geographic ranges simulated for Supplemental Figure 6. All summary statistics are plotted in identical order to the simulation/inference results plotted in Supplemental Figure 6.

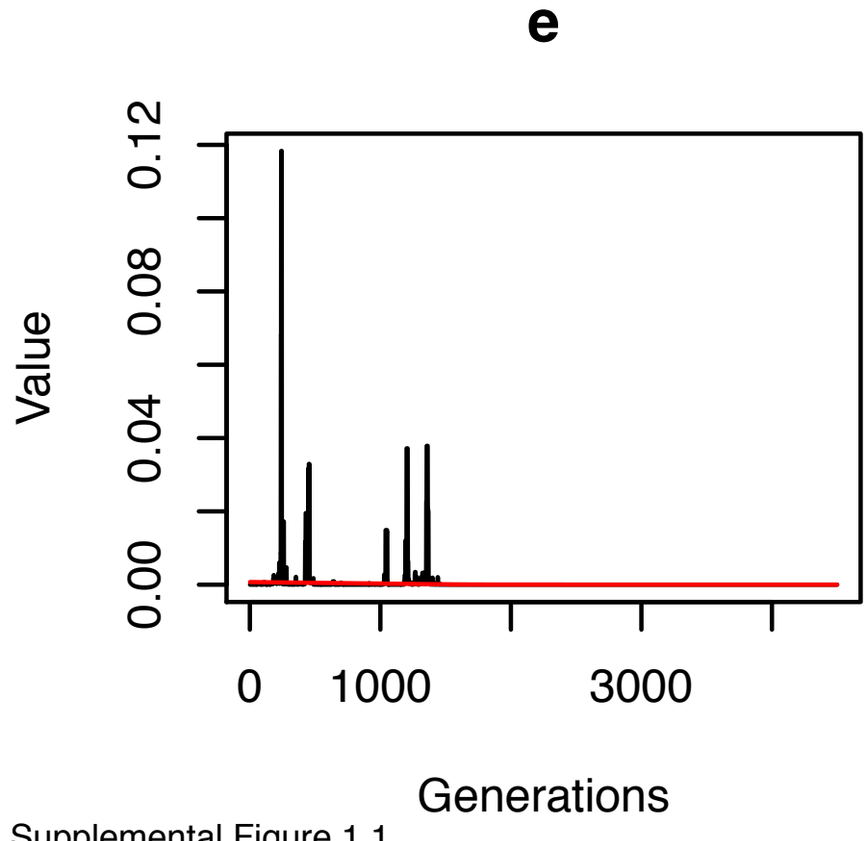
Supplemental Figure 9. ML parameter estimates under DEC and DEC+J for parameters *d* (grey), *e* (white), and *j* (black). For each parameter, results are plotted in the same order as in Figure 5 and Supplemental Table 1. The * represents a bar that has been cut off for display purposes; the true value for this bar is *j*=3 (the taxon is Galapagos *Microlophus*), DEC+J analysis; a value of *j*=3 means that for this taxon, ML inference estimated that founder-event speciation events receive 100% of the probability at cladogenesis, and the other types of events (sympatric-range copying, sympatric-subset, and vicariance) receive 0%.

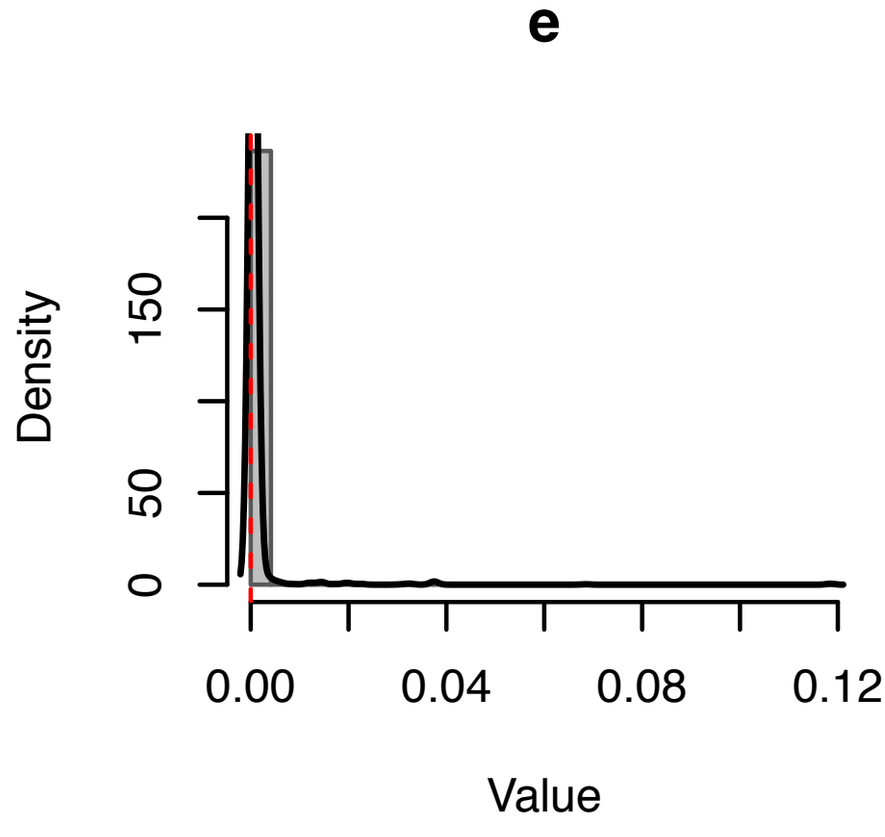
Supplemental Data 1a: Cladogenesis calculations. This Excel file demonstrates the cladogenesis calculations that occur at speciation events in the DEC and DEC+J models, and the parameterization of cladogenesis done in BioGeoBEARS. The different worksheets shown the matrix in terms of parameter symbols, weights, and resulting probabilities. Cladogenesis transition matrices are shown only for ancestral areas of size 2, 3, and 4; beyond 4 areas, the size of the matrix is such that graphical depiction is impractical.

Supplemental Data 1b: PDFs of ancestral range estimates. The PDFs in this zipfile are named by clade name and constraint abbreviation, corresponding to Table 2. These represent the raw output from BioGeoBEARS, and may be regenerated from scratch using the scripts in the *examples_new* directory at Dryad.

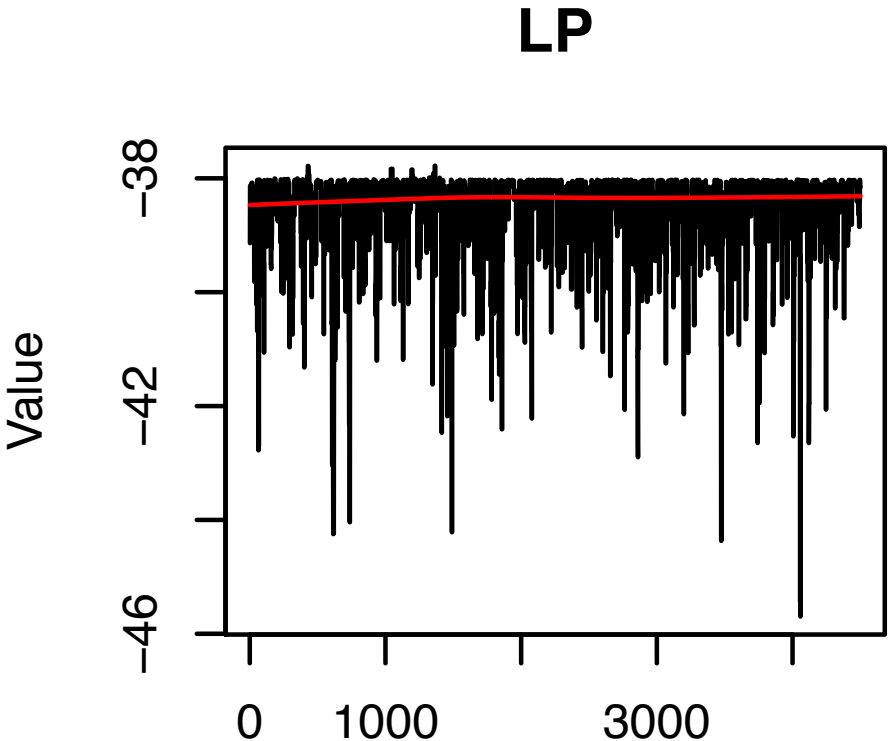


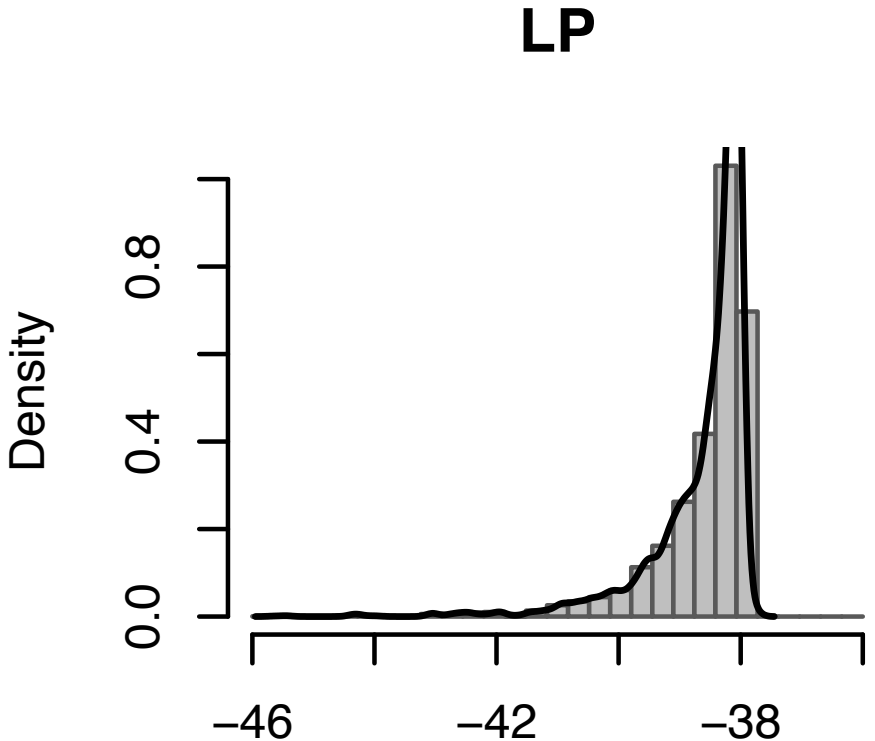




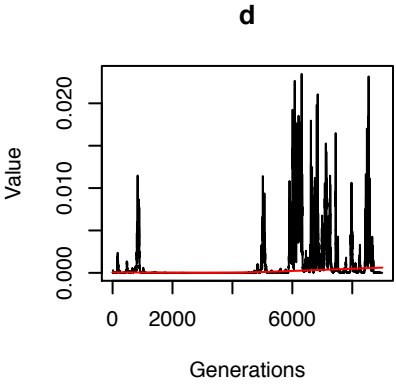


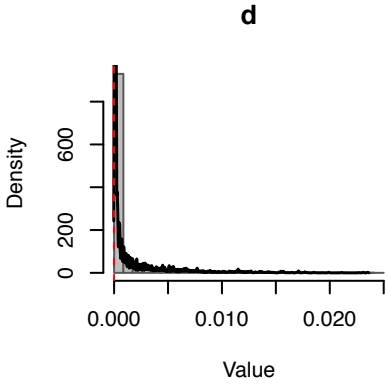
Supplemental Figure 1.1

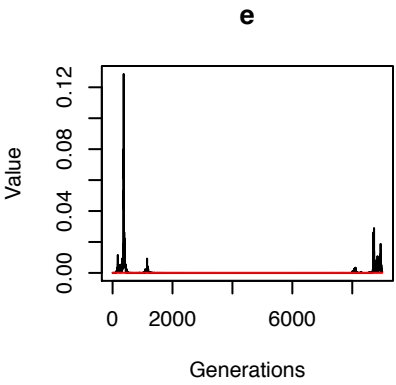


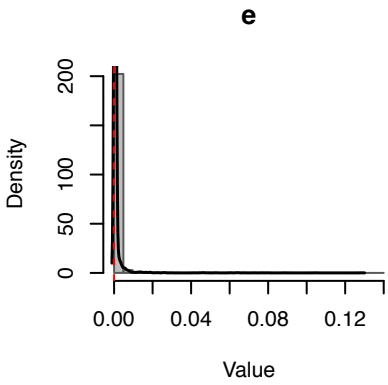


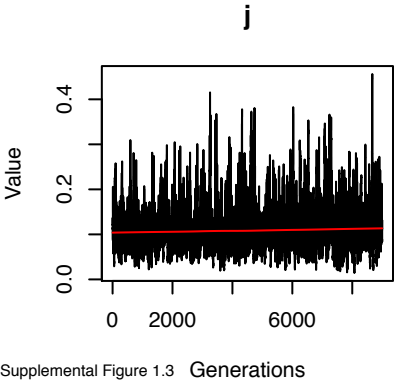
Supplemental Figure 1.2

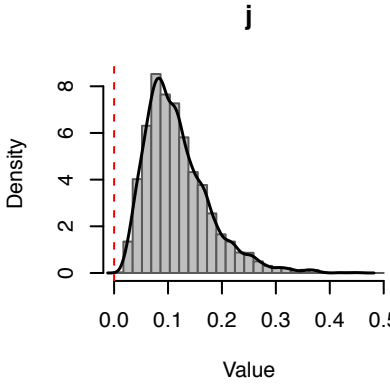




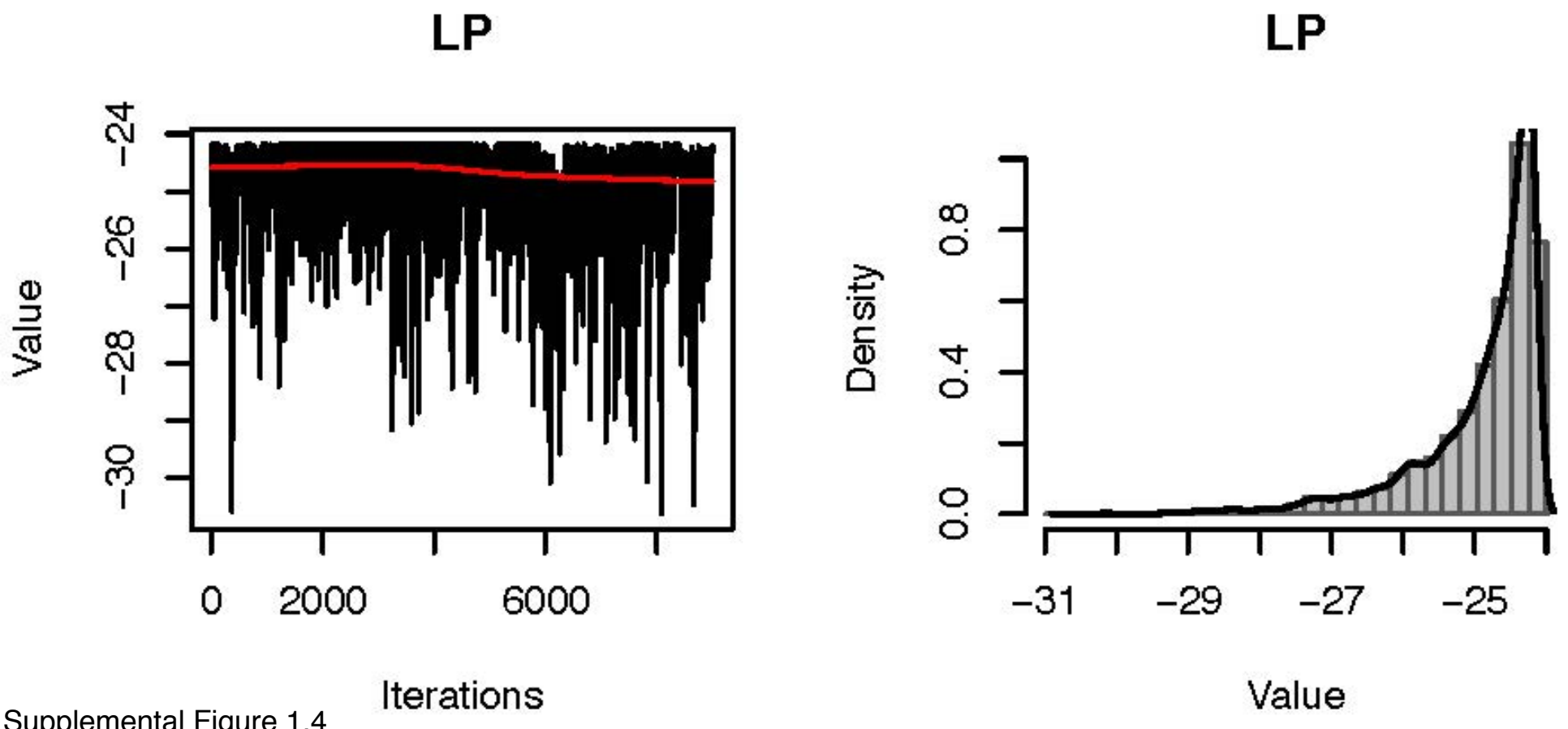




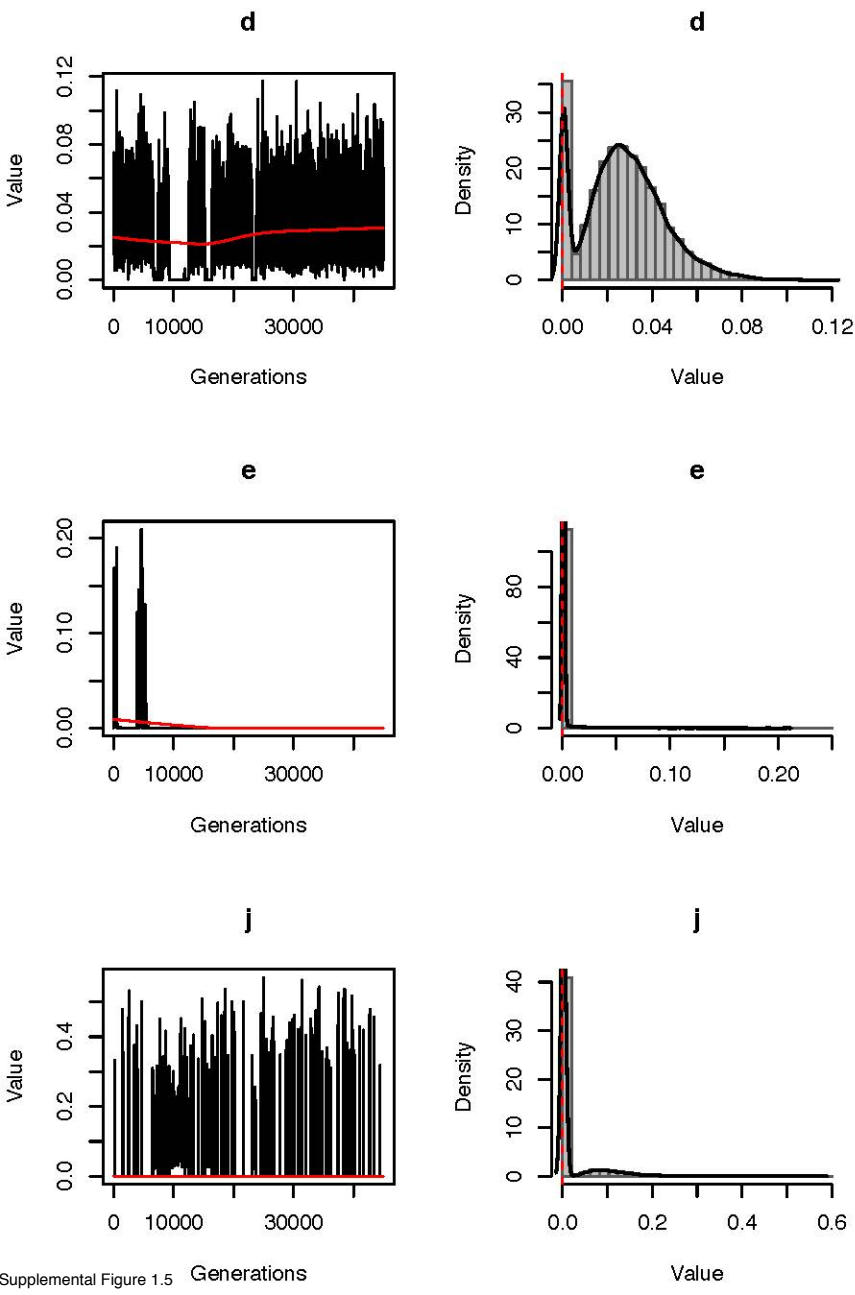




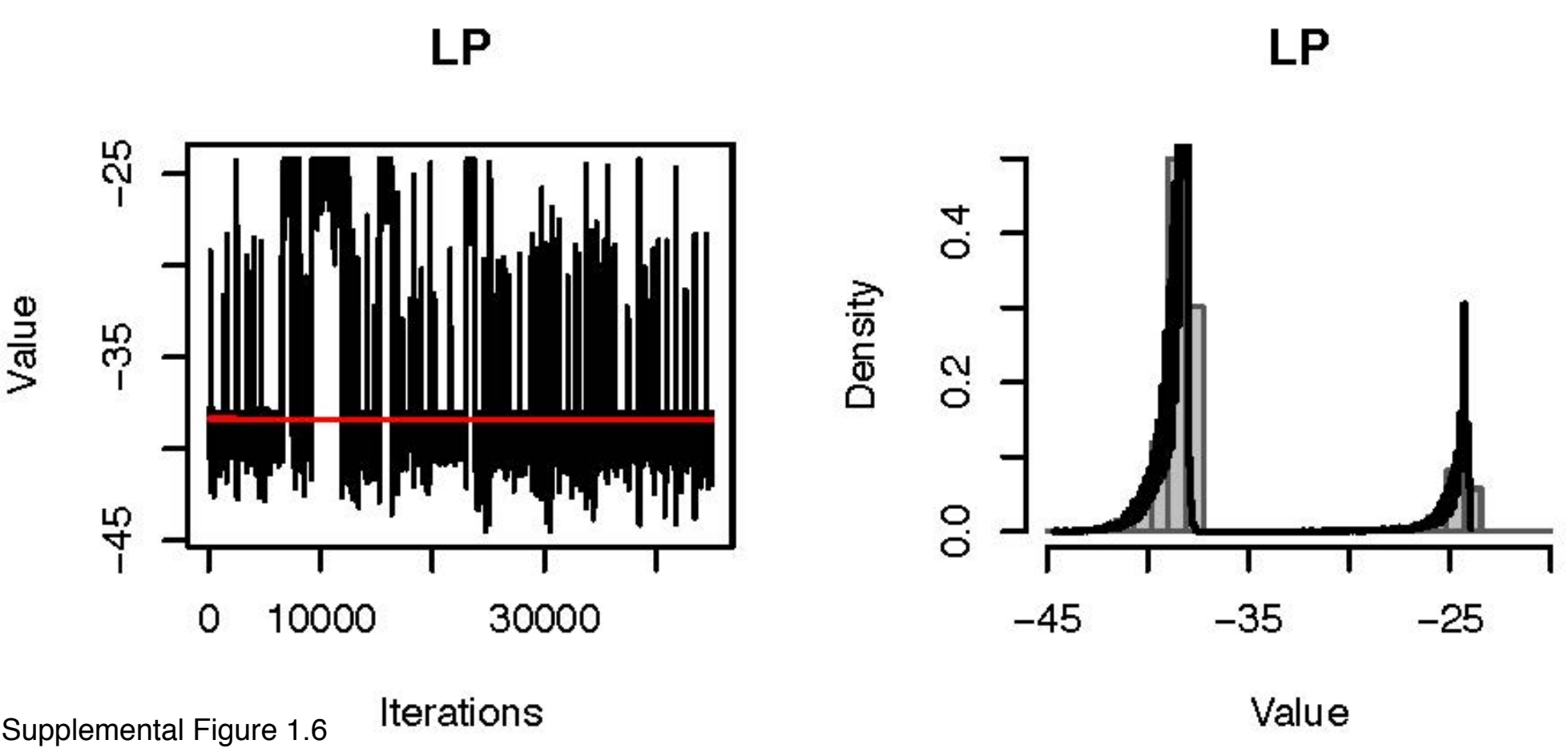
Supplemental Figure 1.3



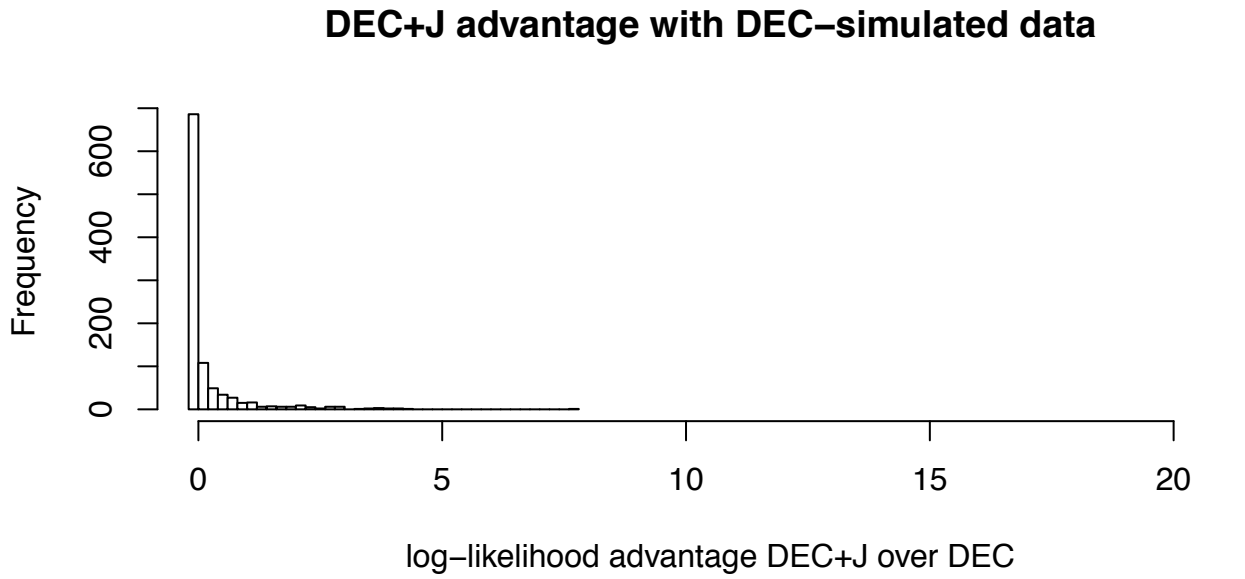
Supplemental Figure 1.4

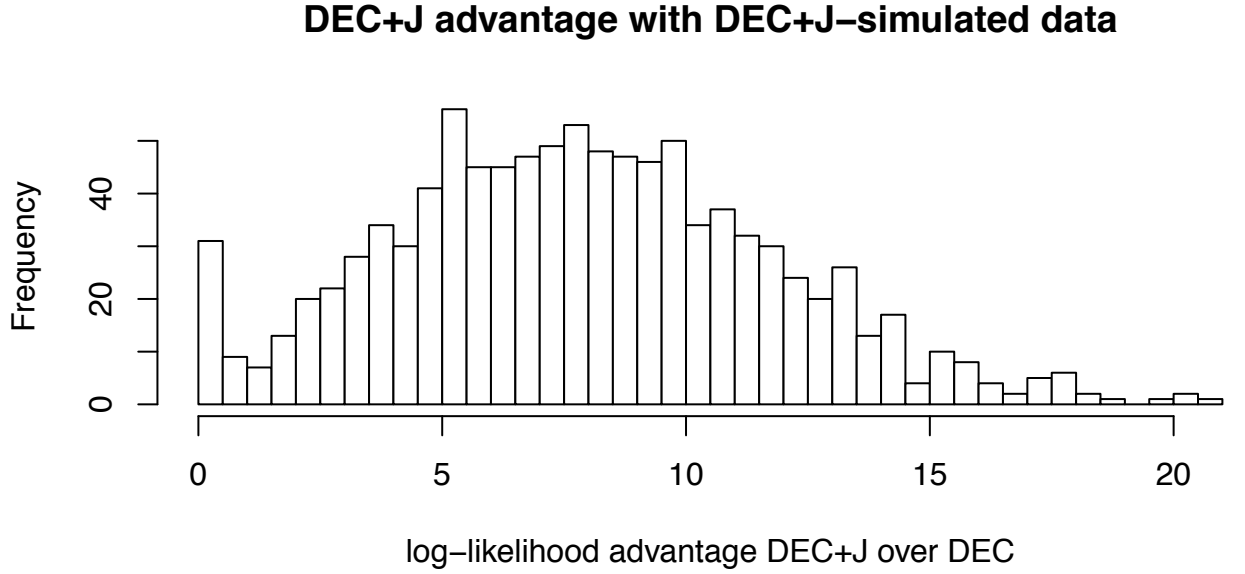


Supplemental Figure 1.5

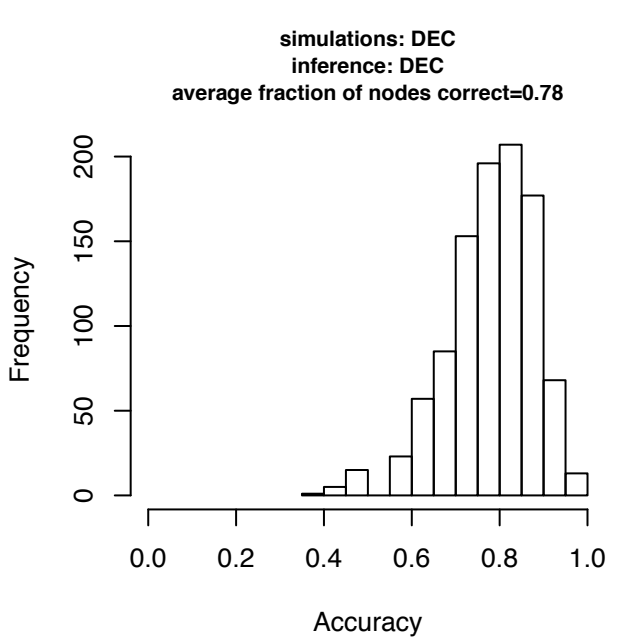


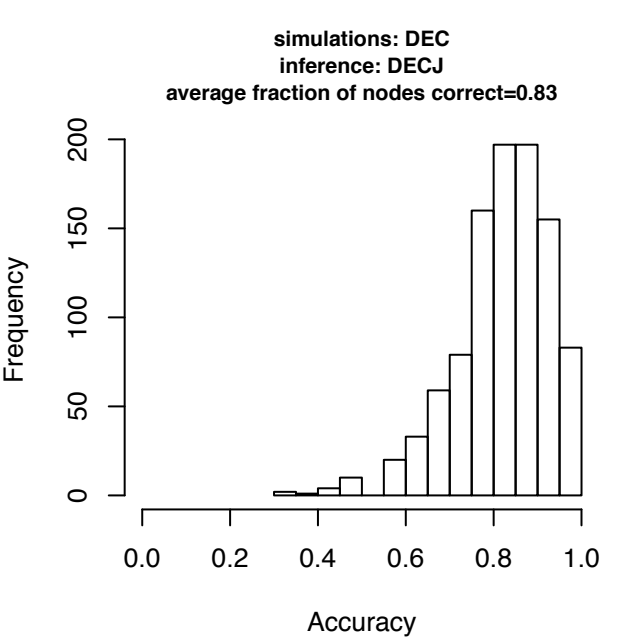
Supplemental Figure 1.6

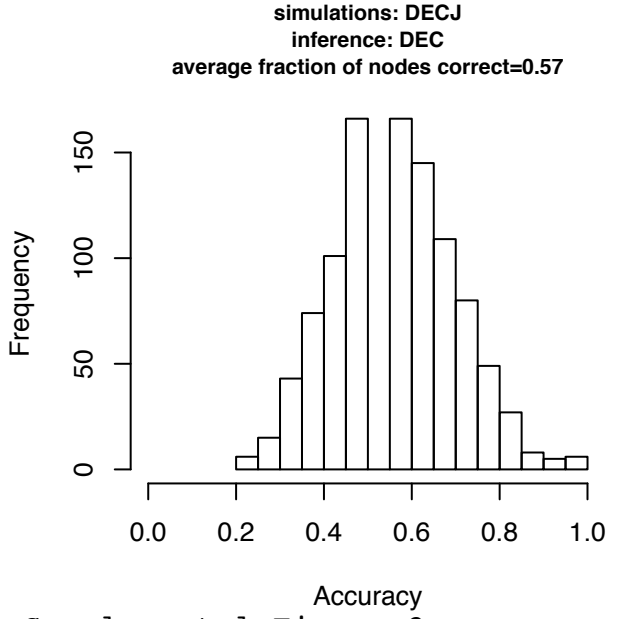


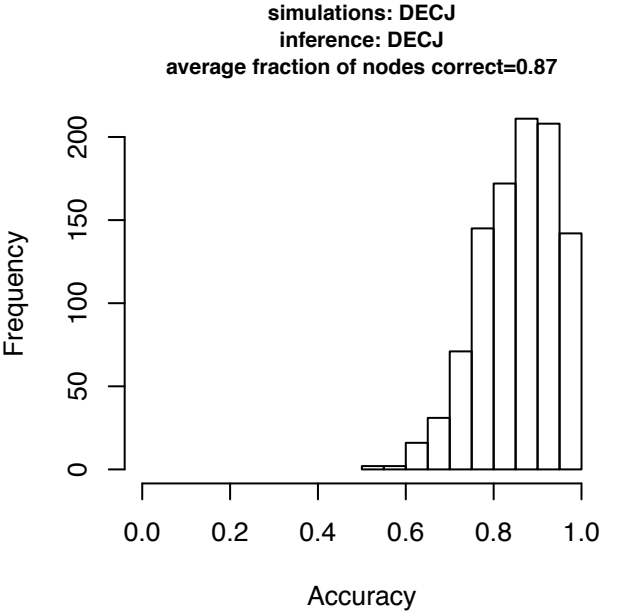


Supplemental Figure 2

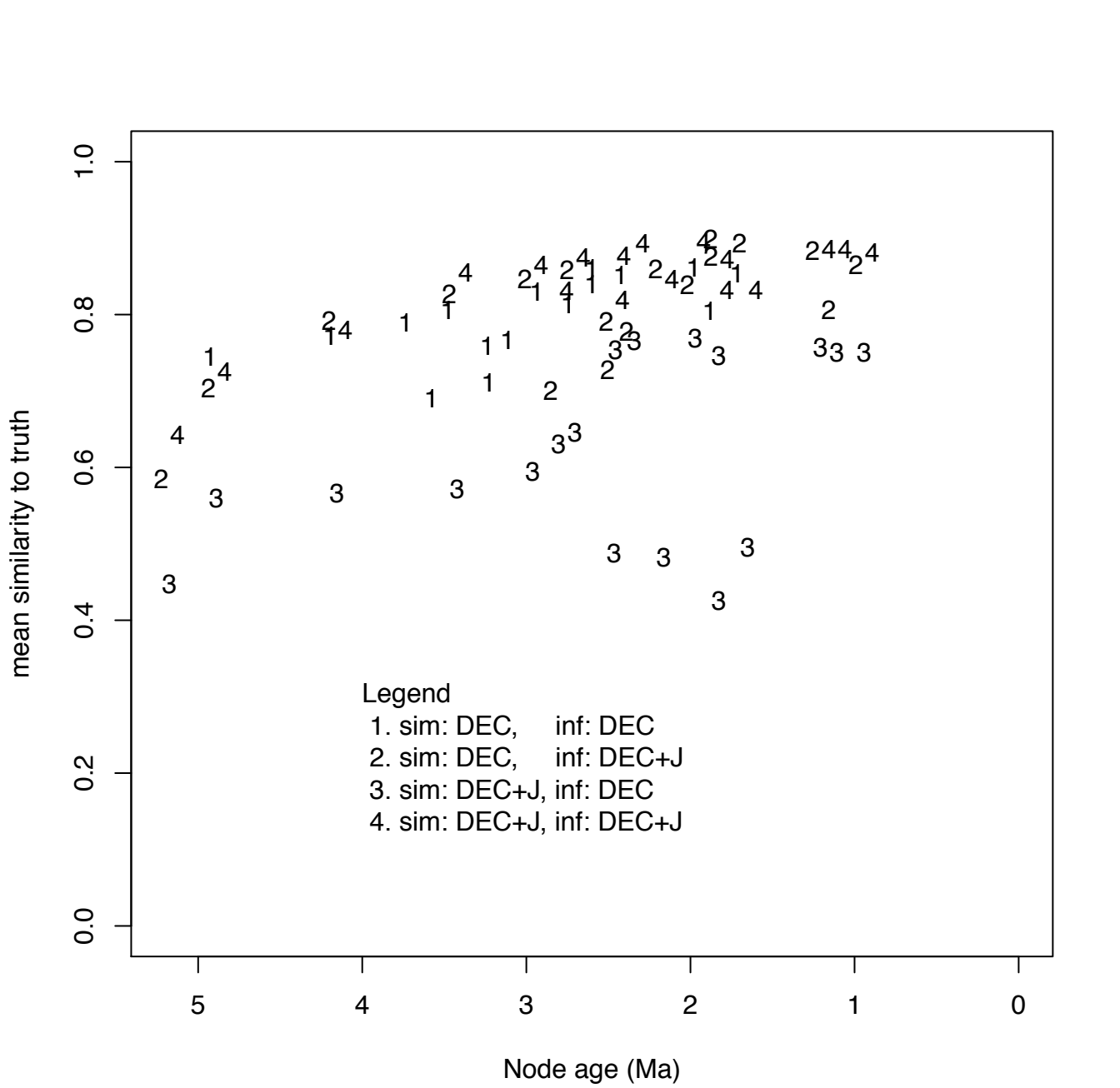




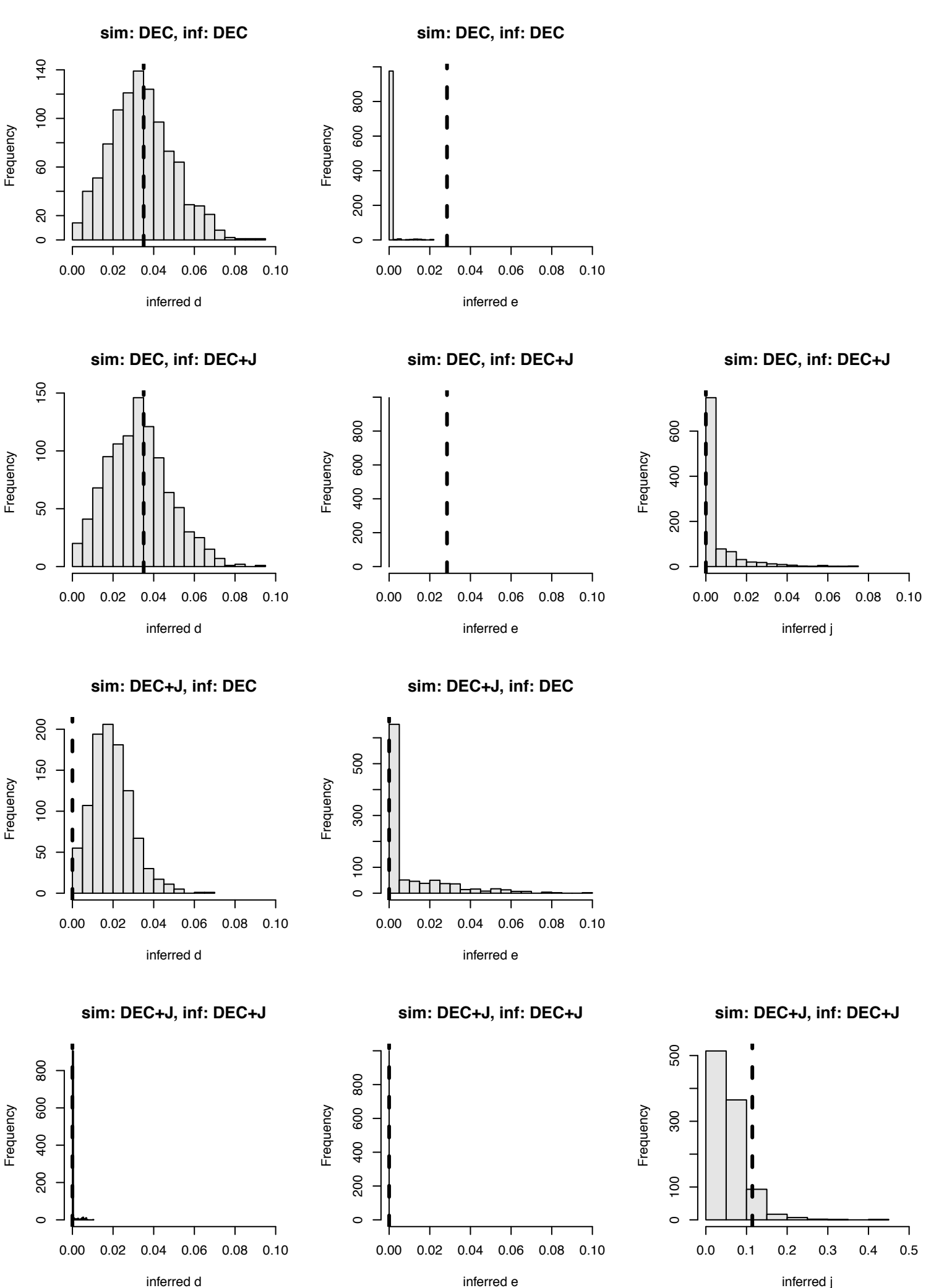




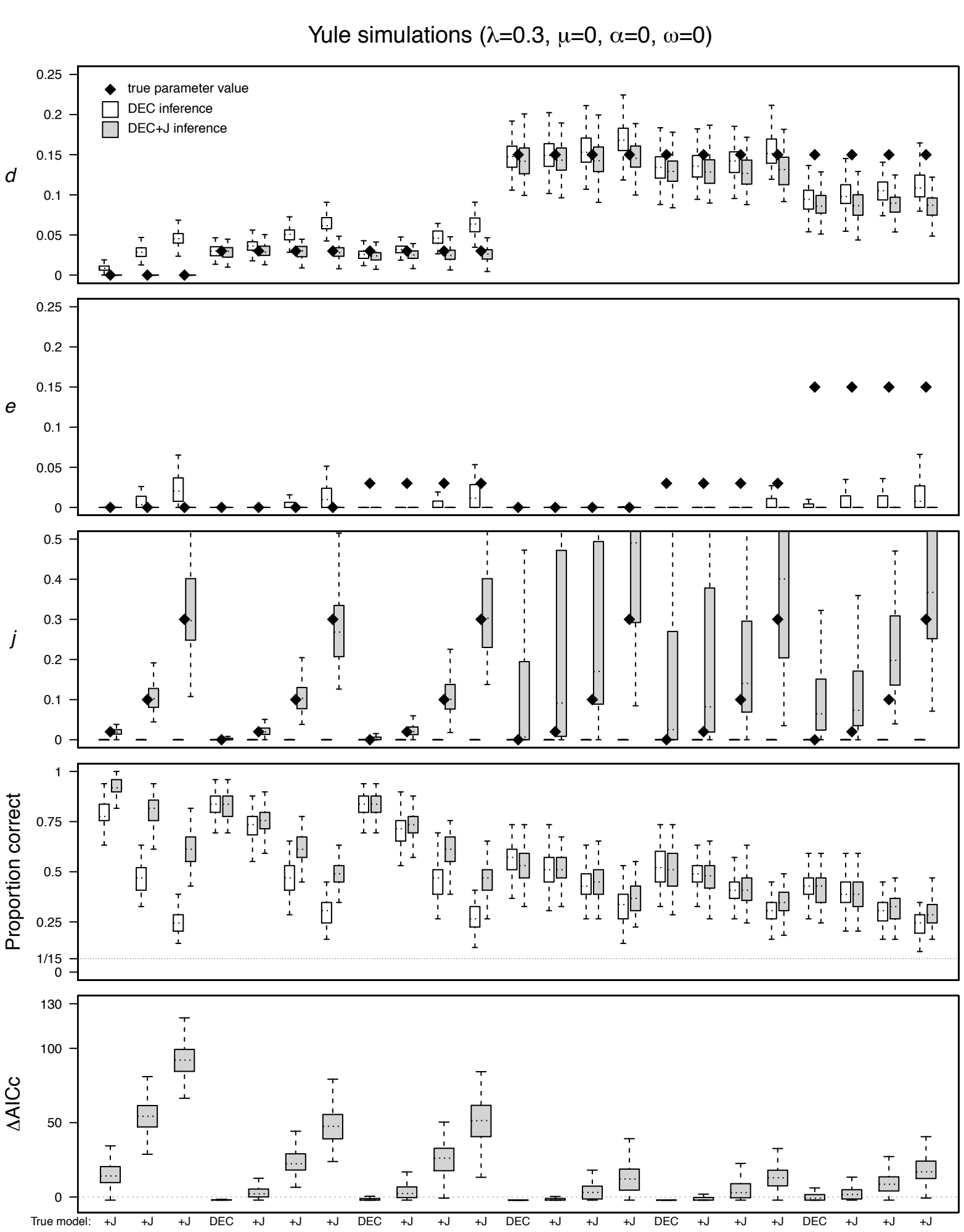
Supplemental Figure 3



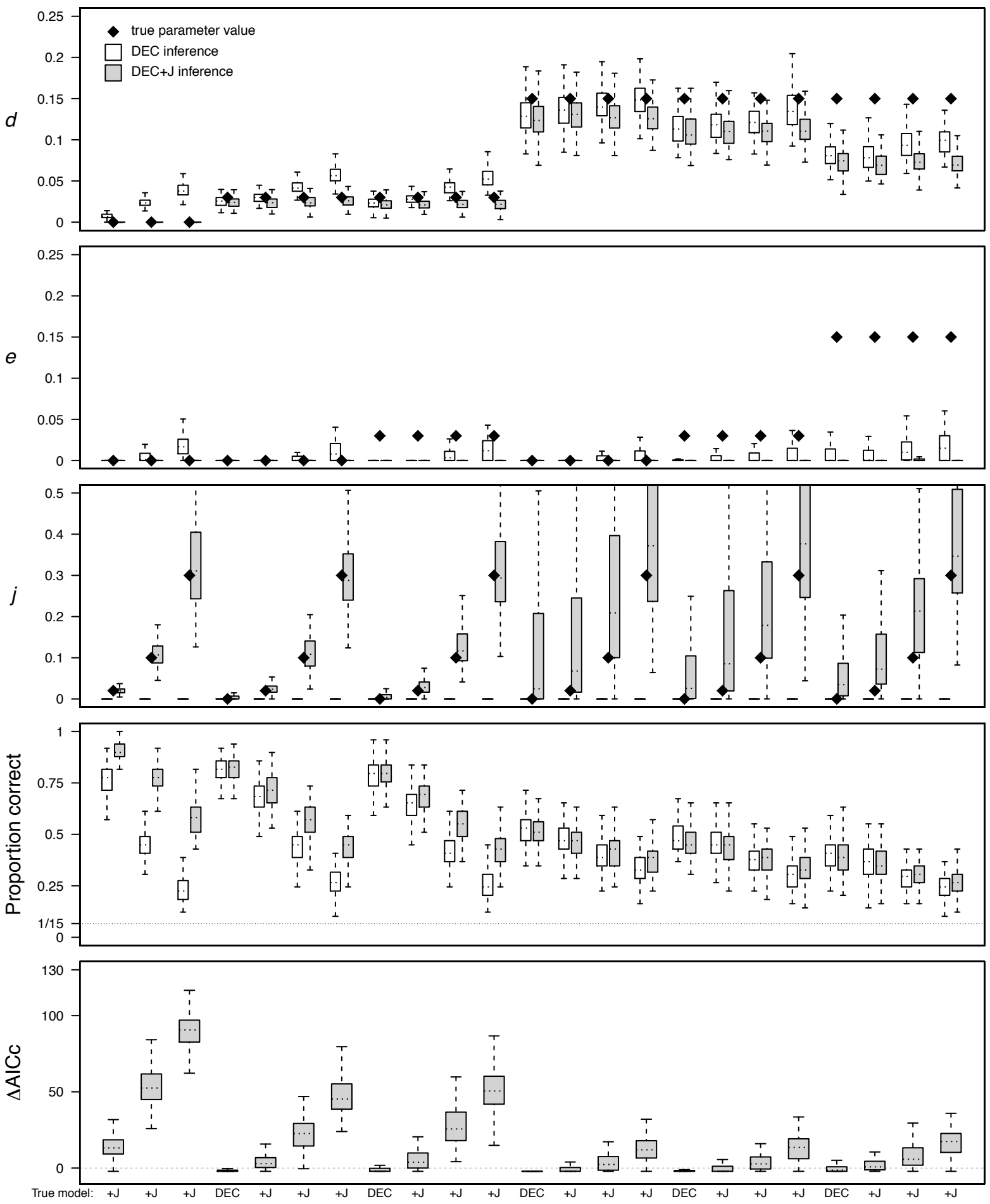
Supplemental Figure 4



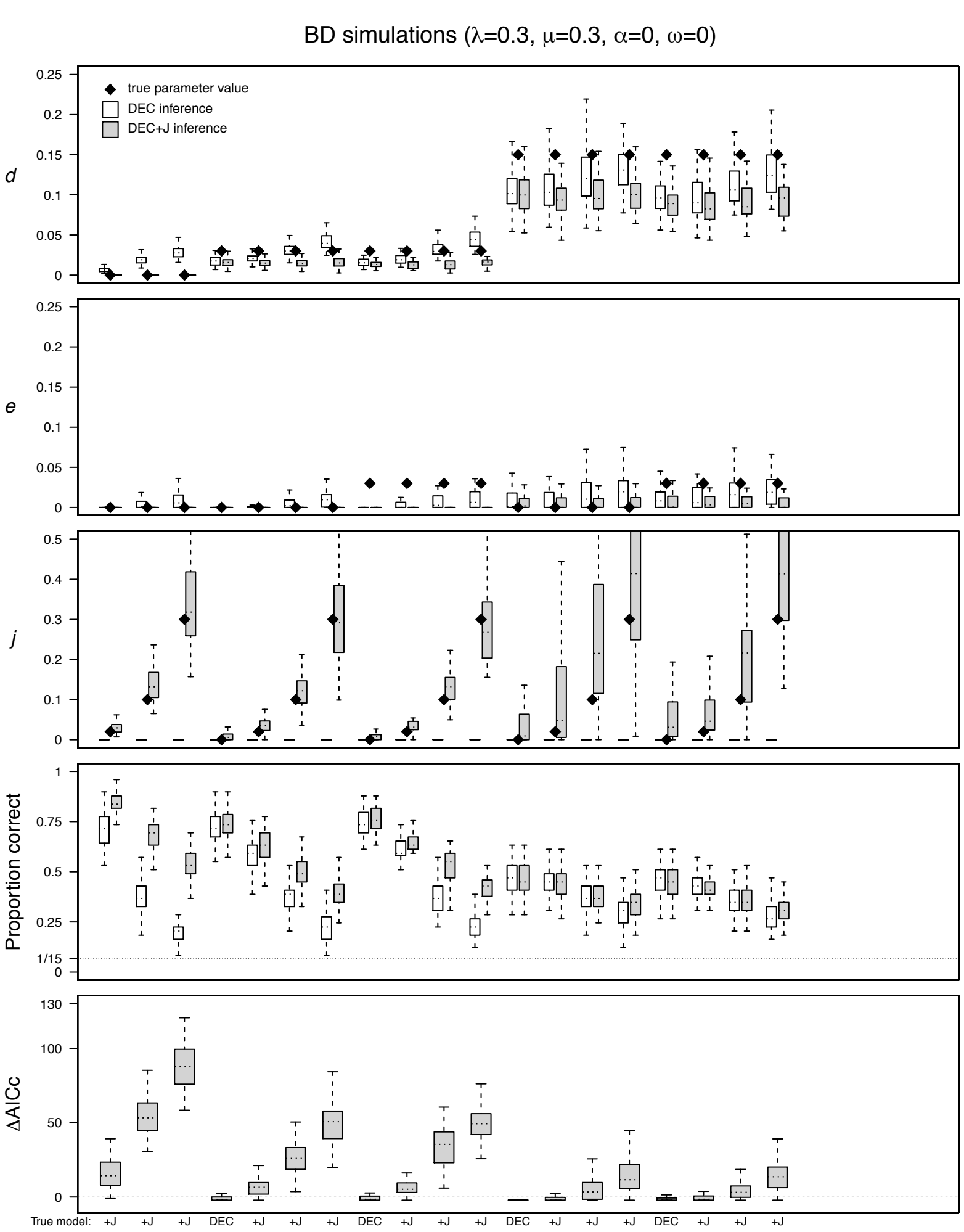
Supplemental Figure 5



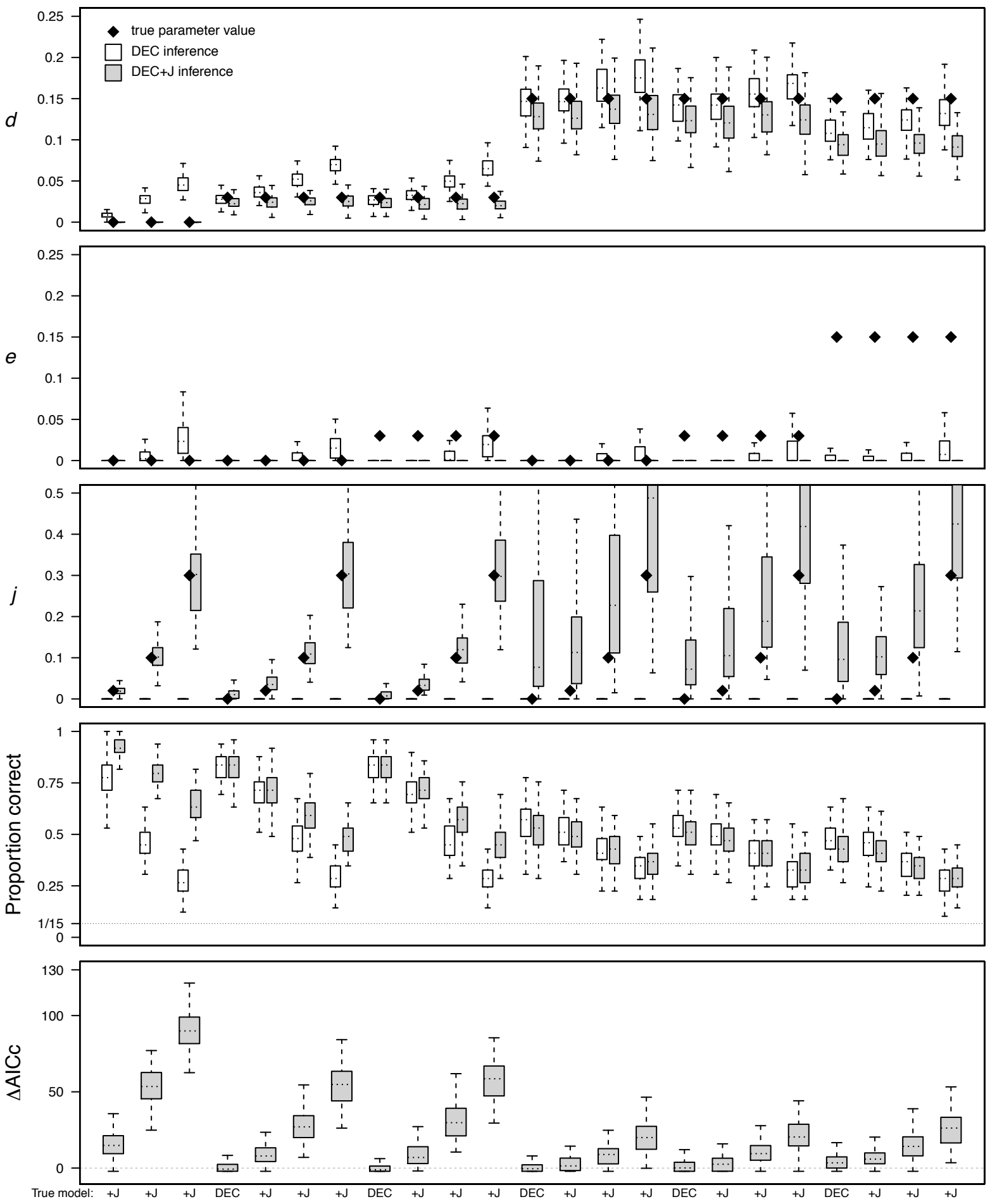
Supplemental Figure 6.1



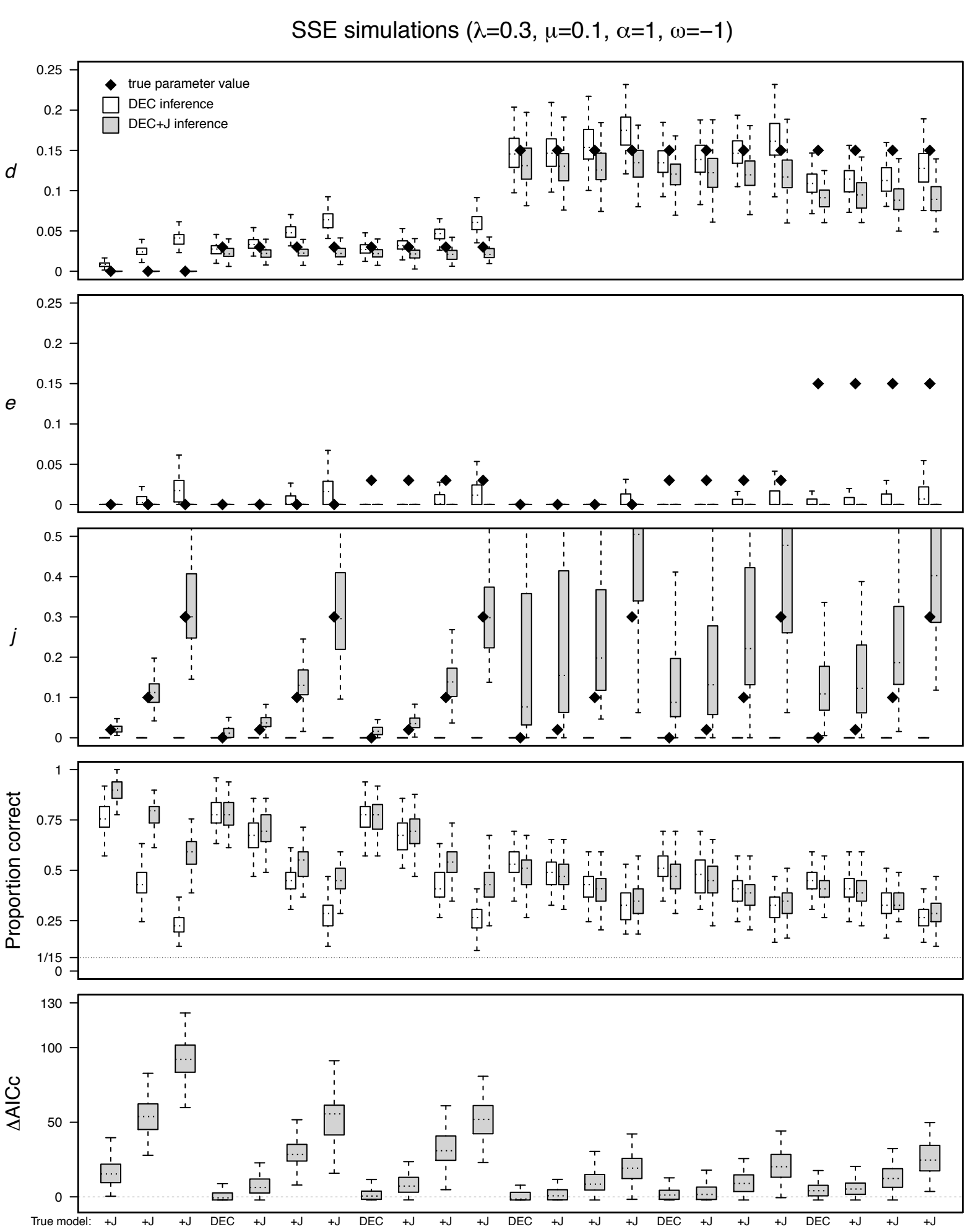
Supplemental Figure 6.2



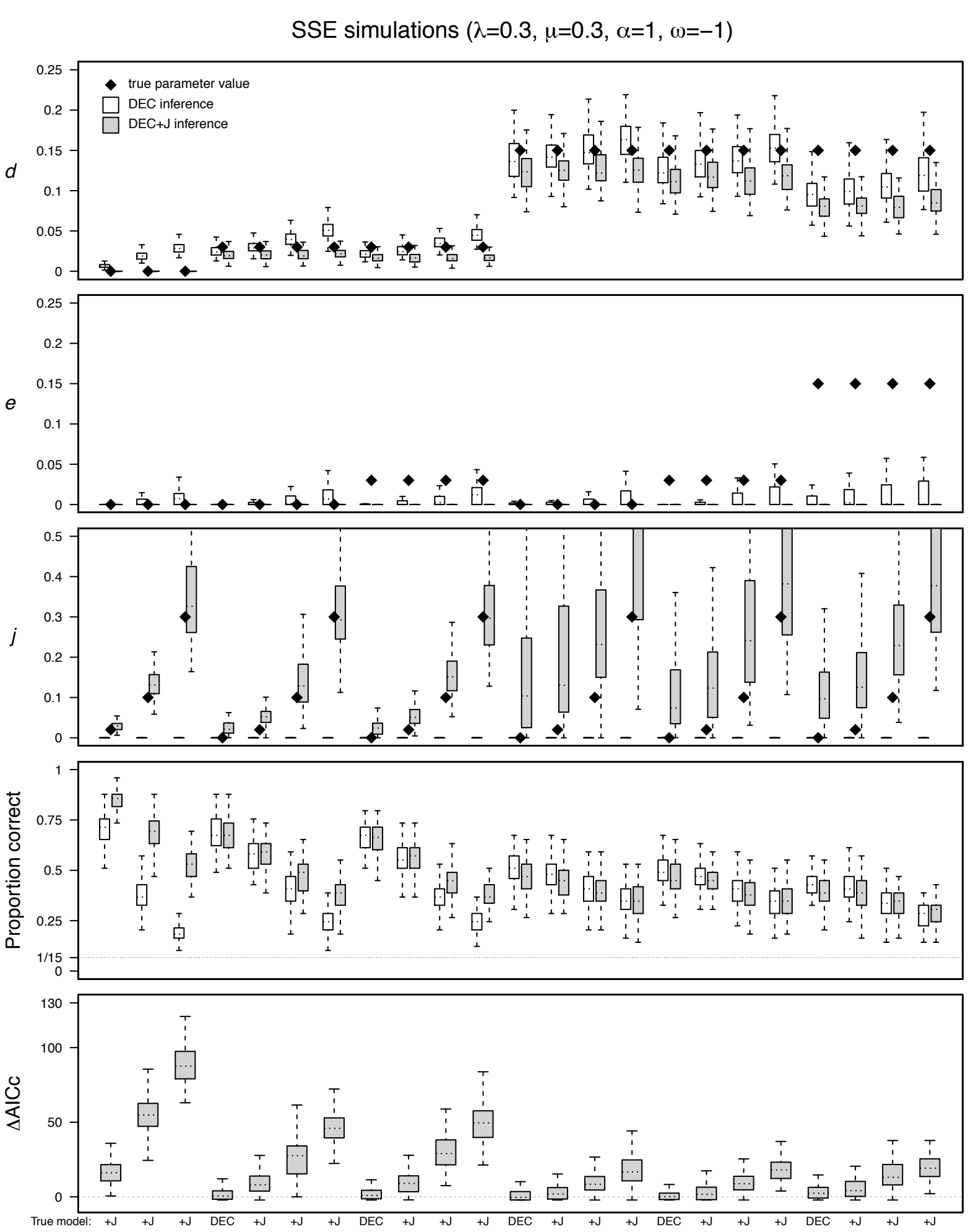
Supplemental Figure 6.3



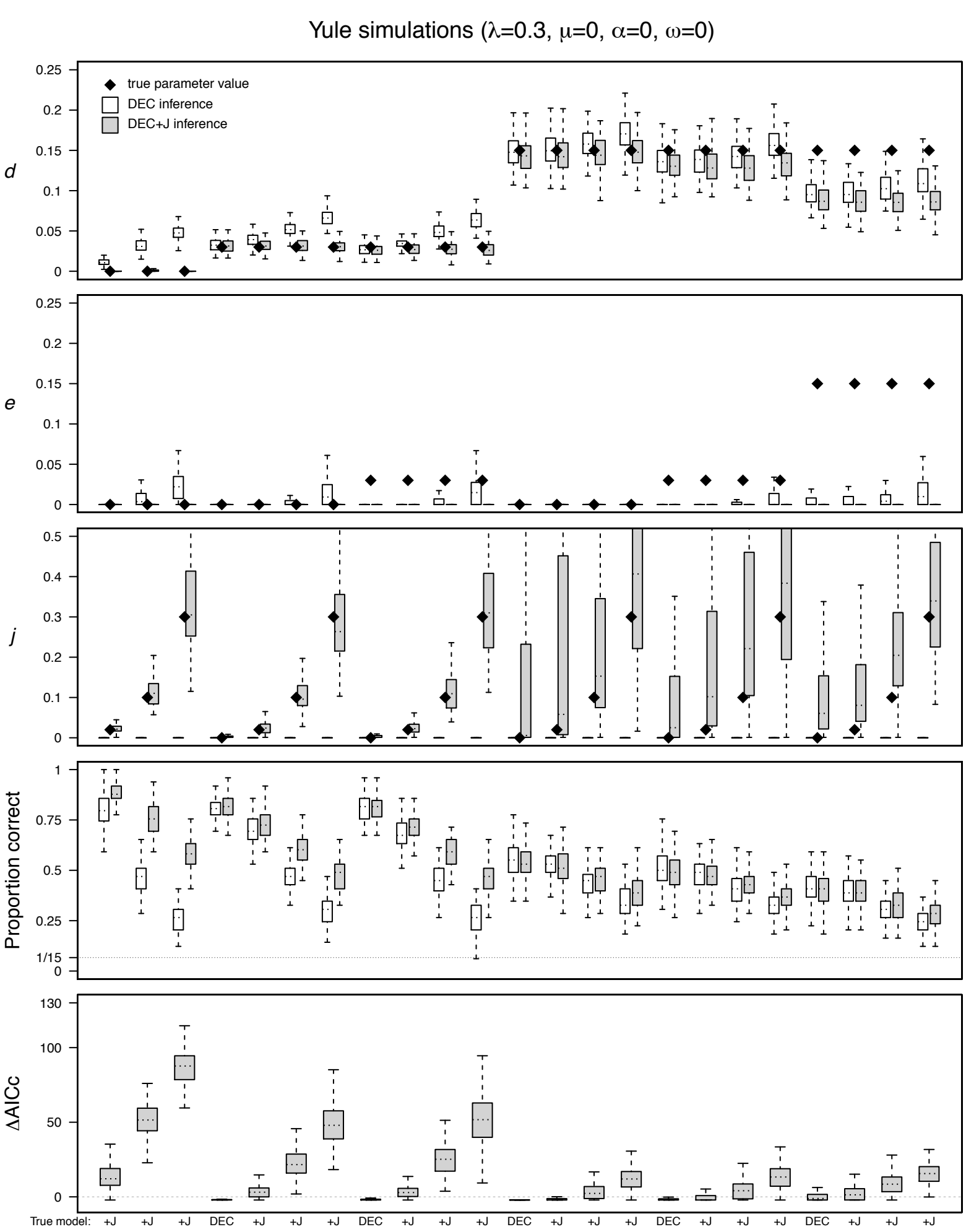
Supplemental Figure 6.4



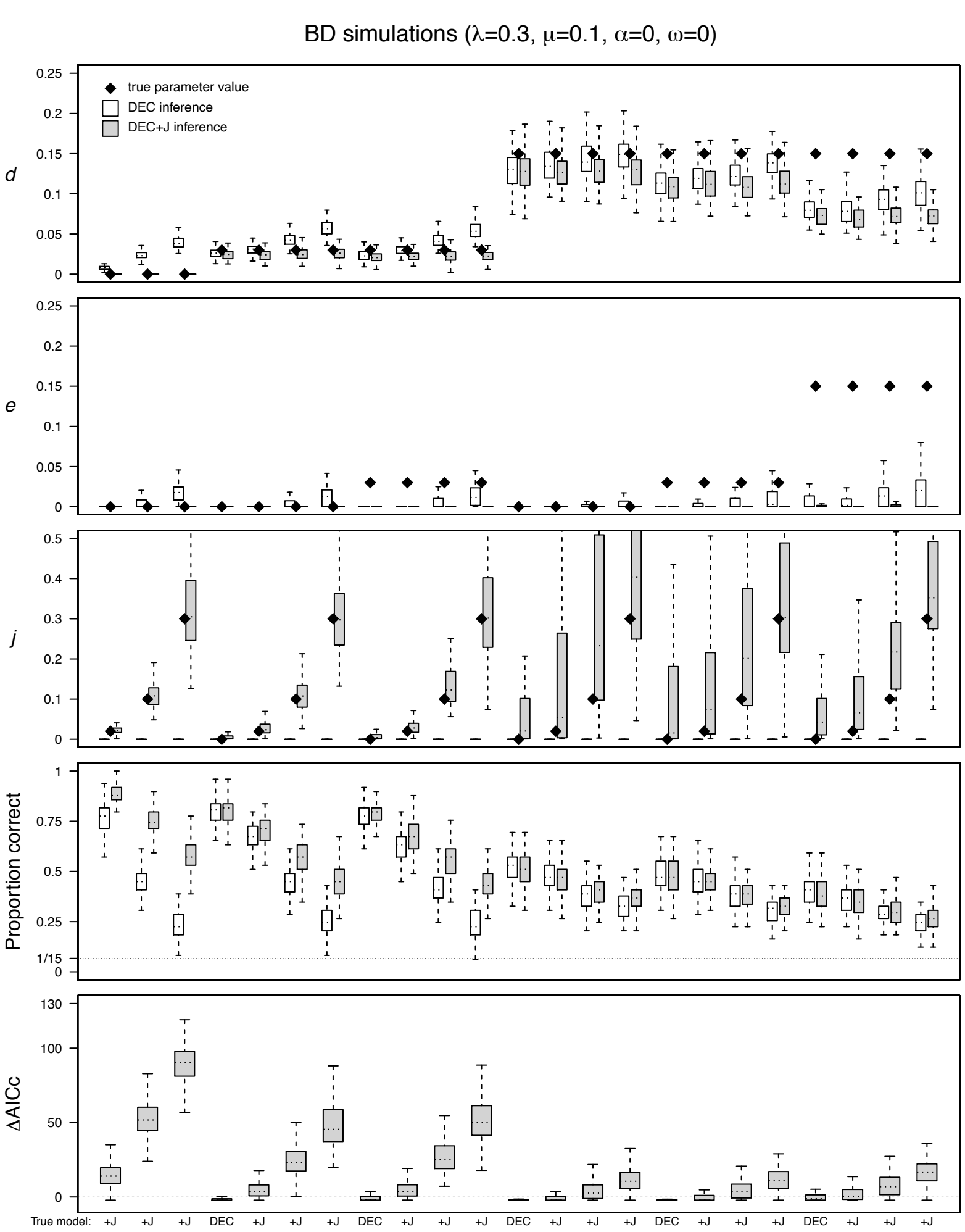
Supplemental Figure 6.5



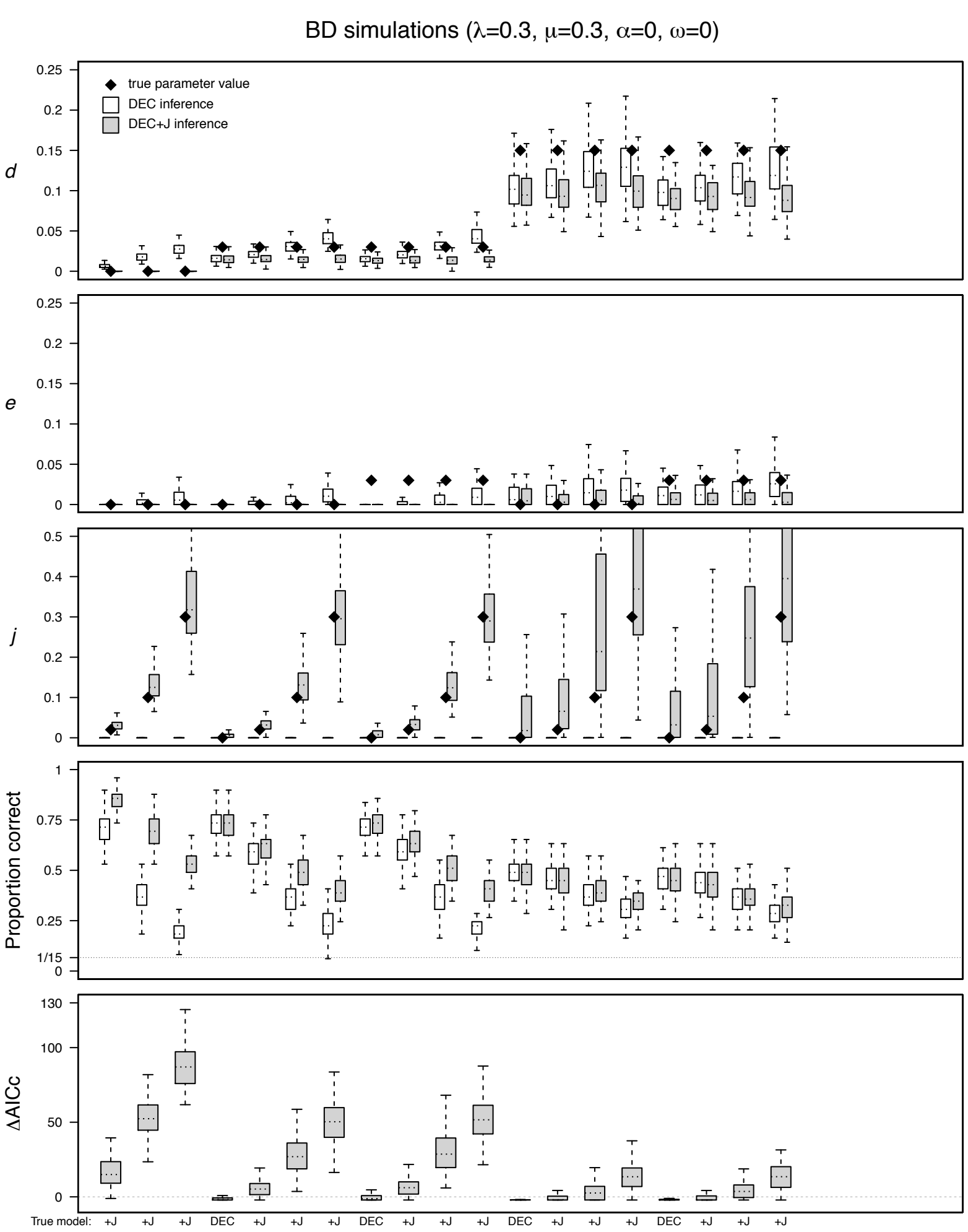
Supplemental Figure 6.6



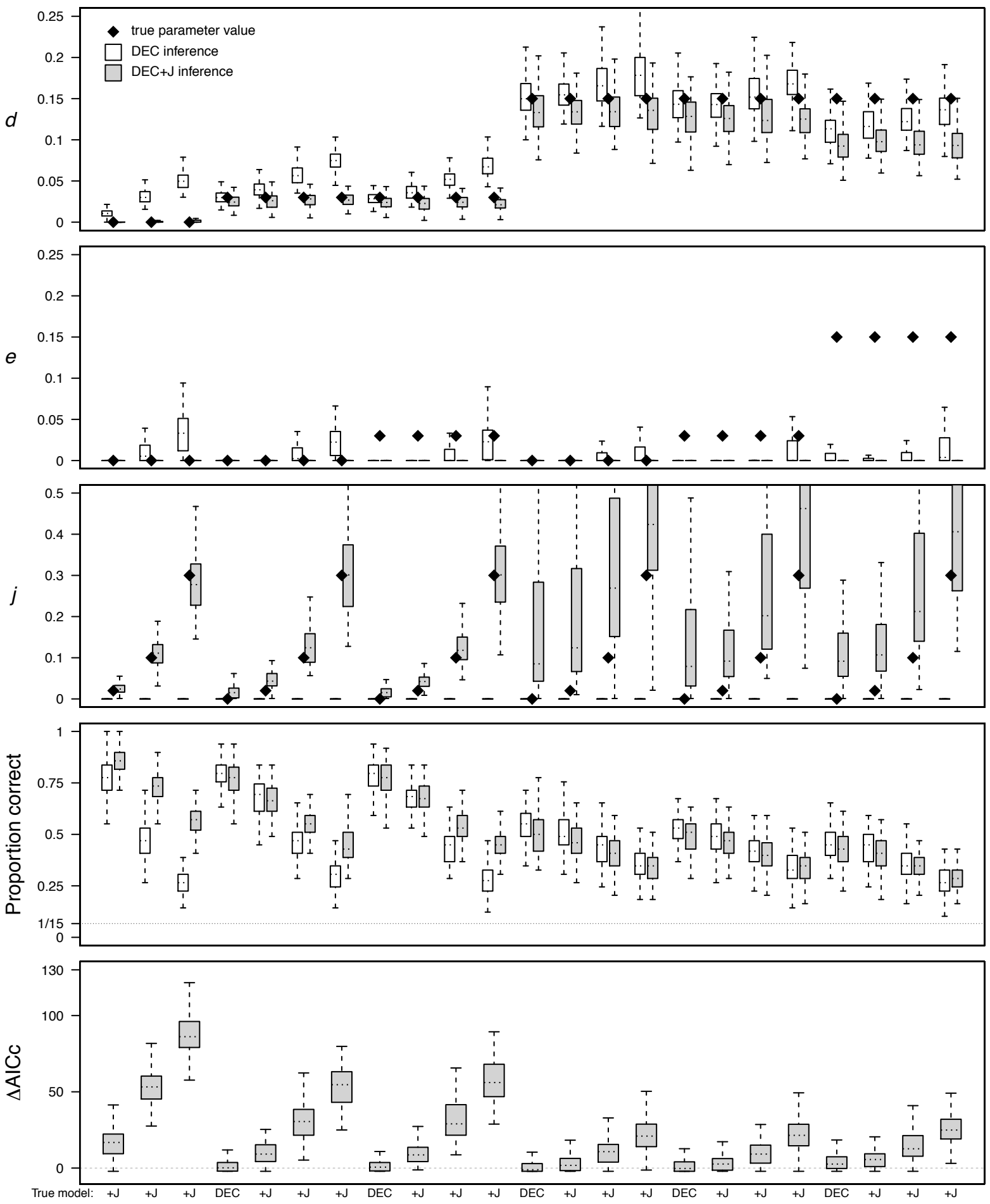
Supplemental Figure 7.1



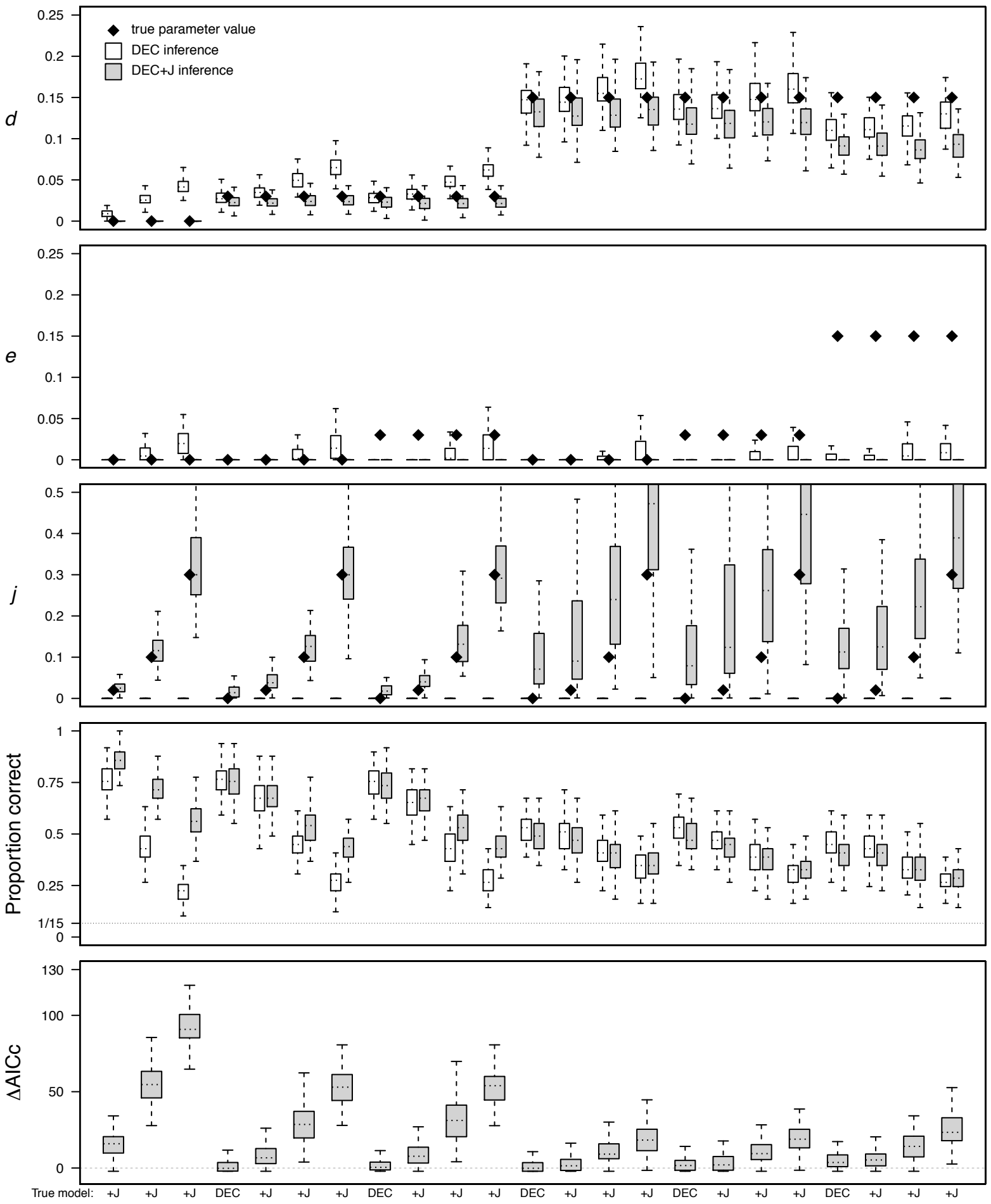
Supplemental Figure 7.2



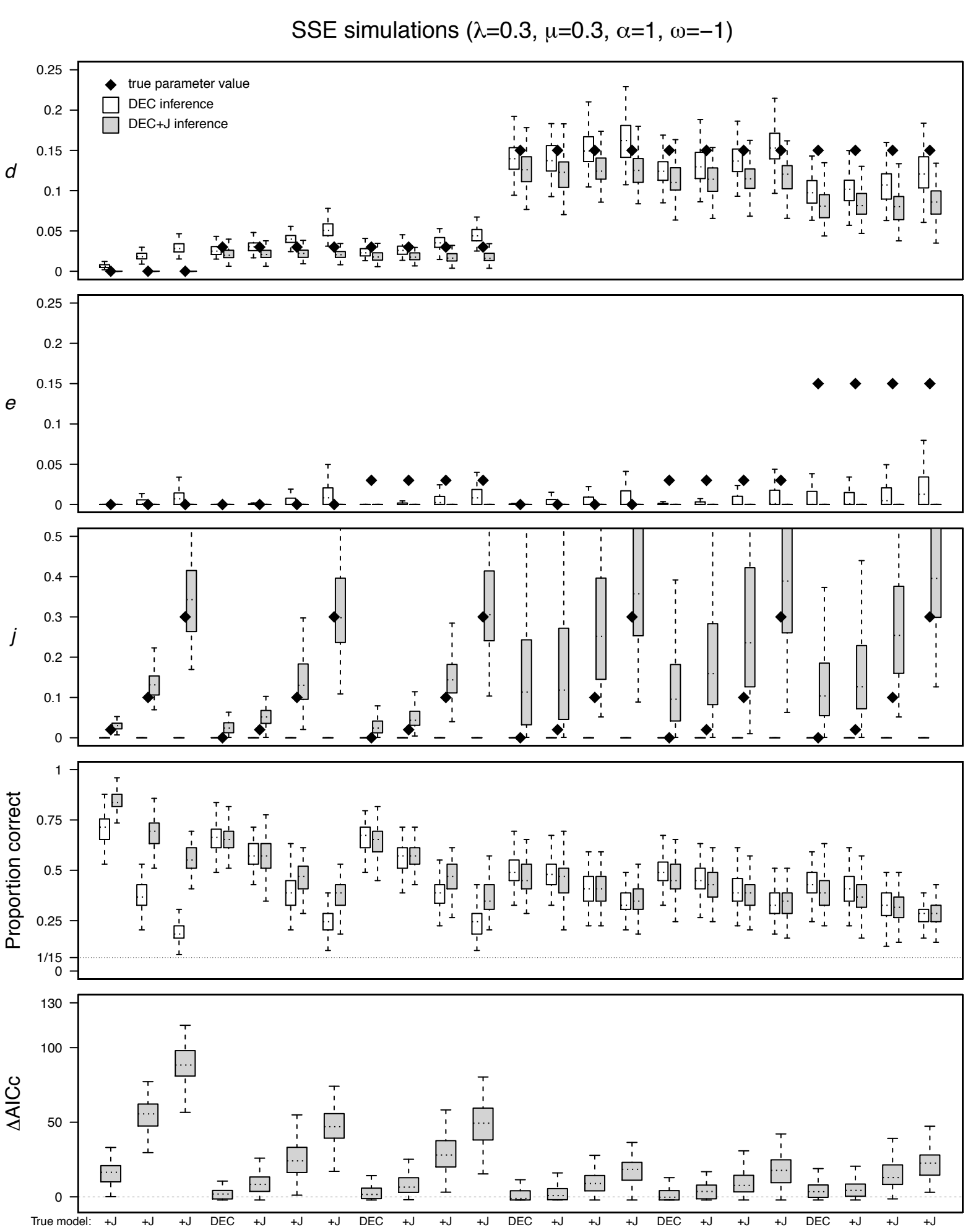
Supplemental Figure 7.3



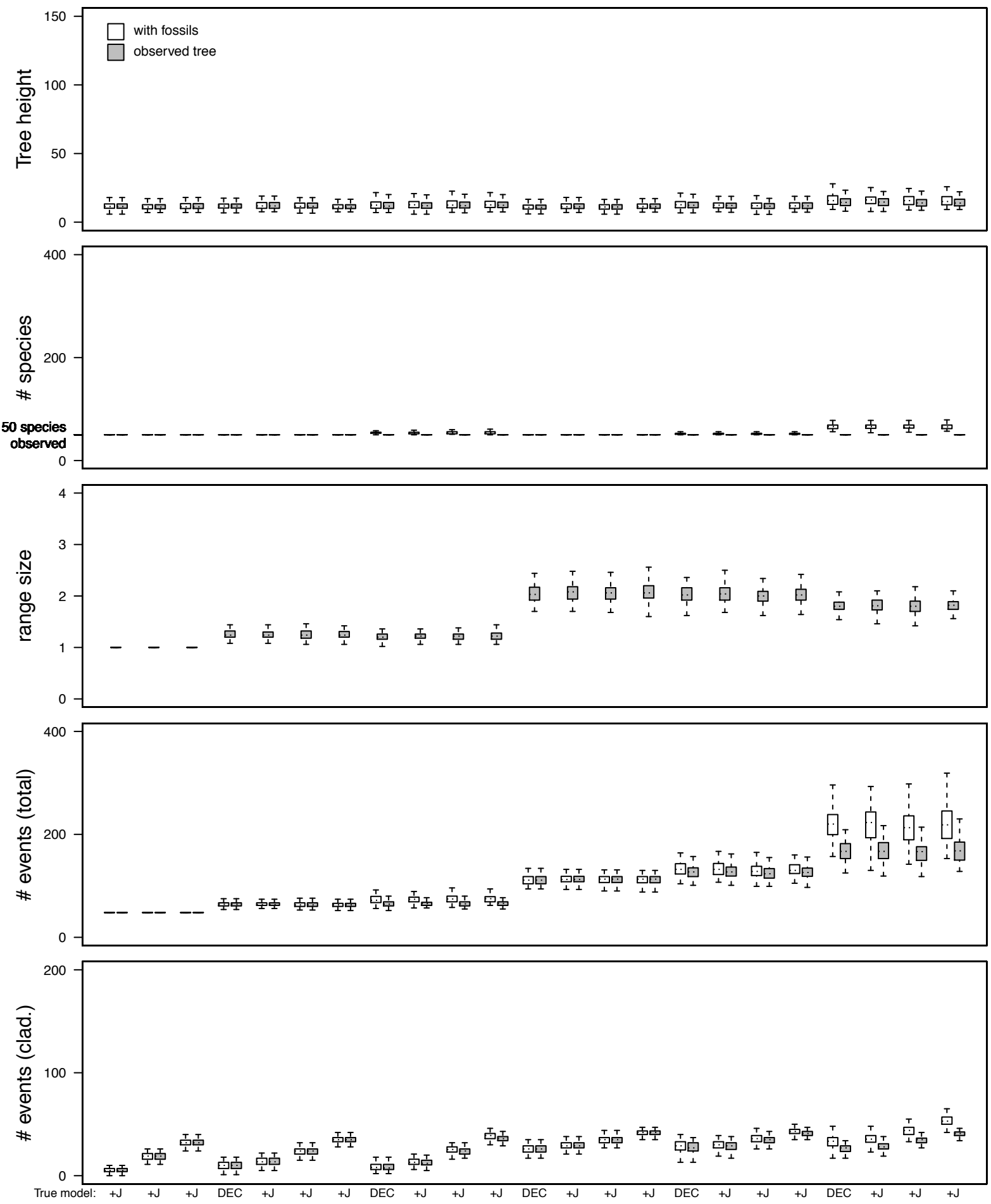
Supplemental Figure 7.4



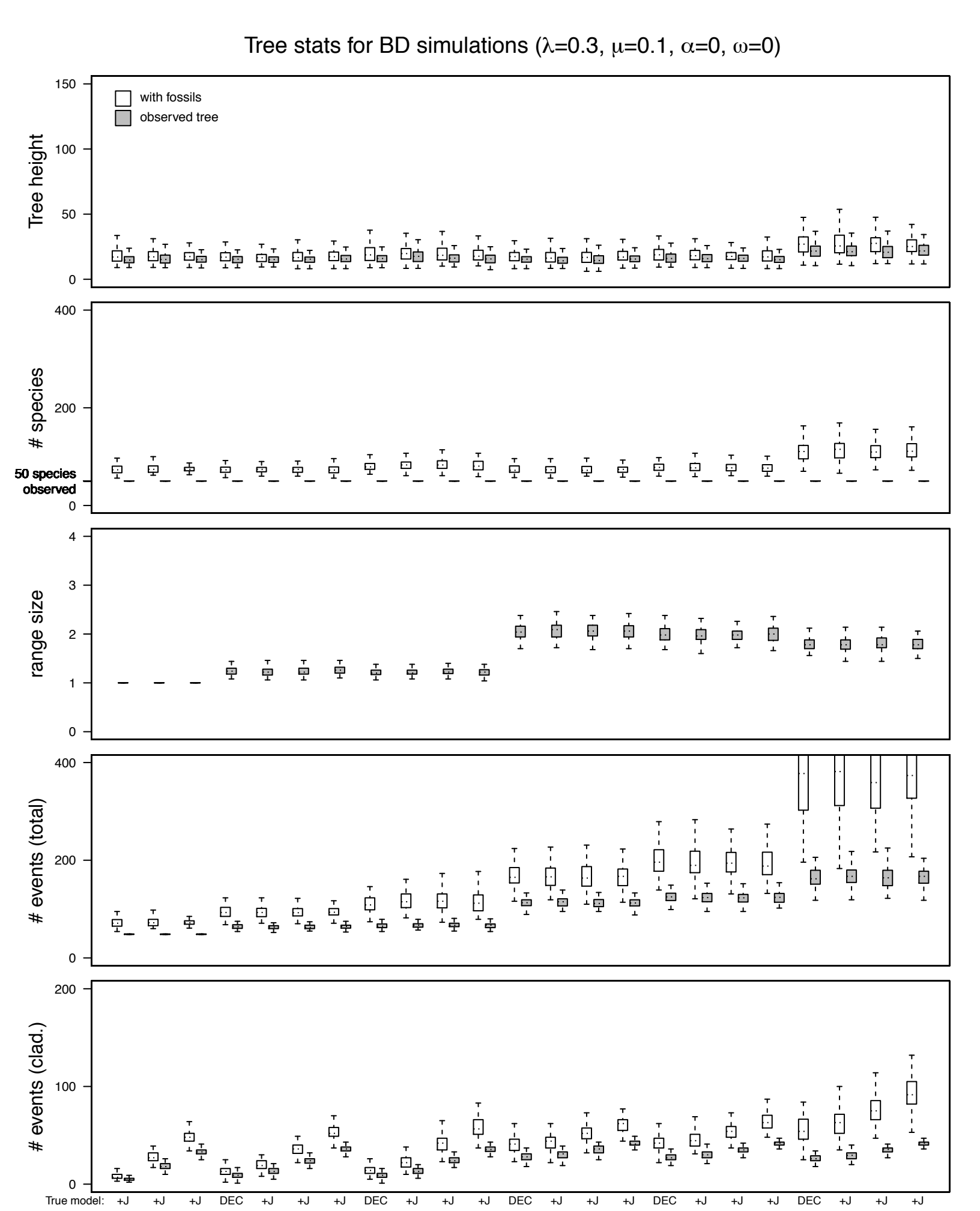
Supplemental Figure 7.5



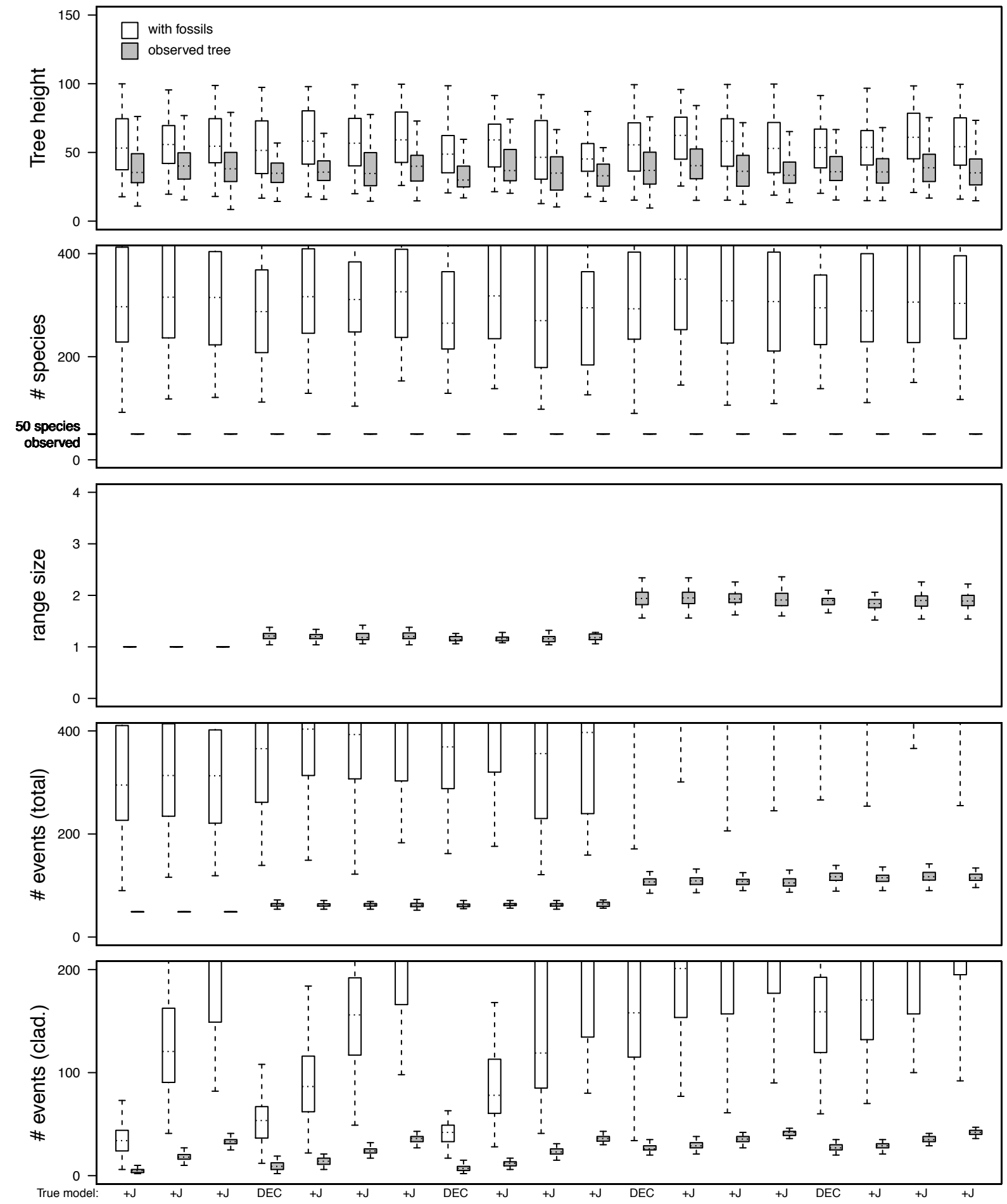
Supplemental Figure 7.6



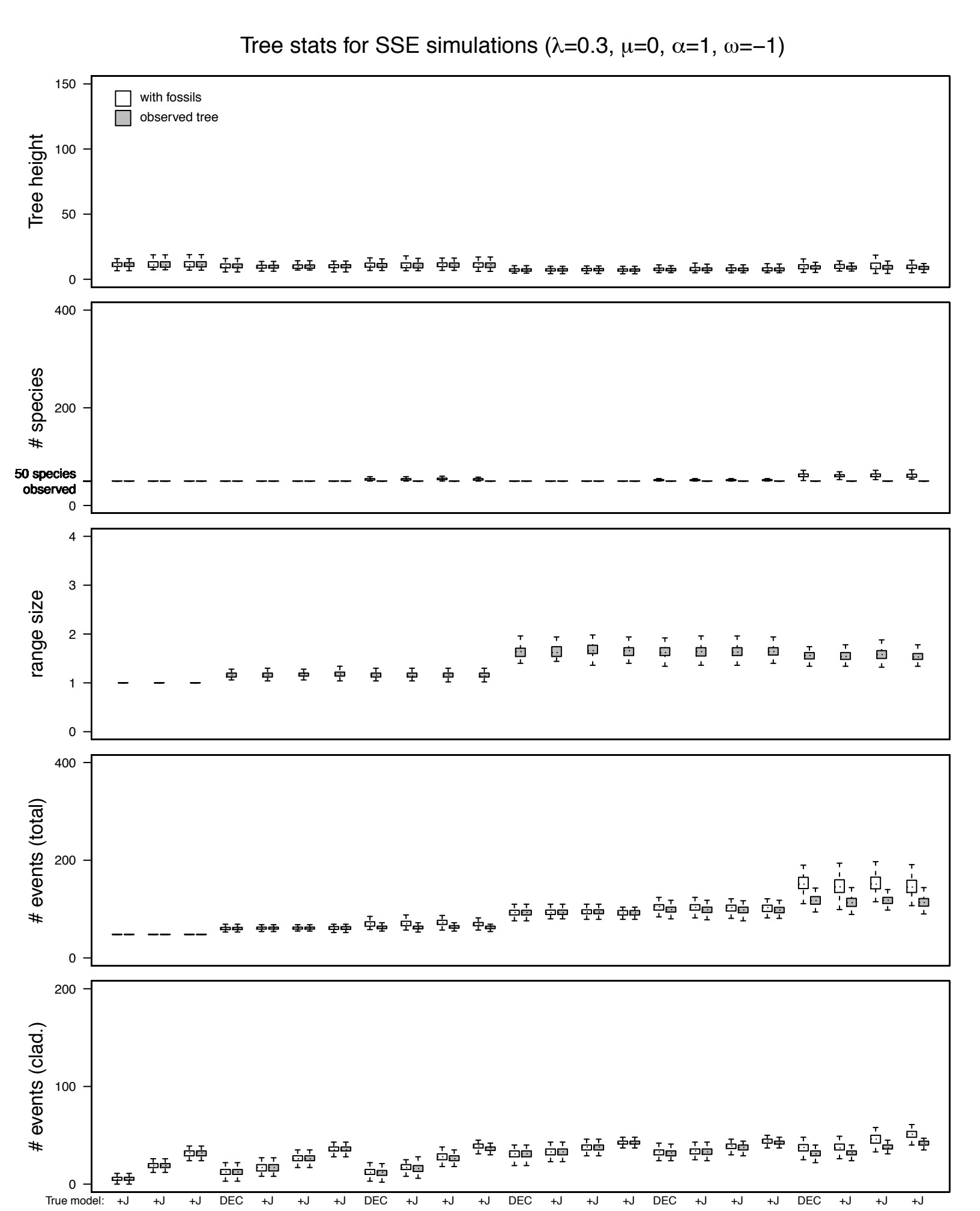
Supplemental Figure 8.1



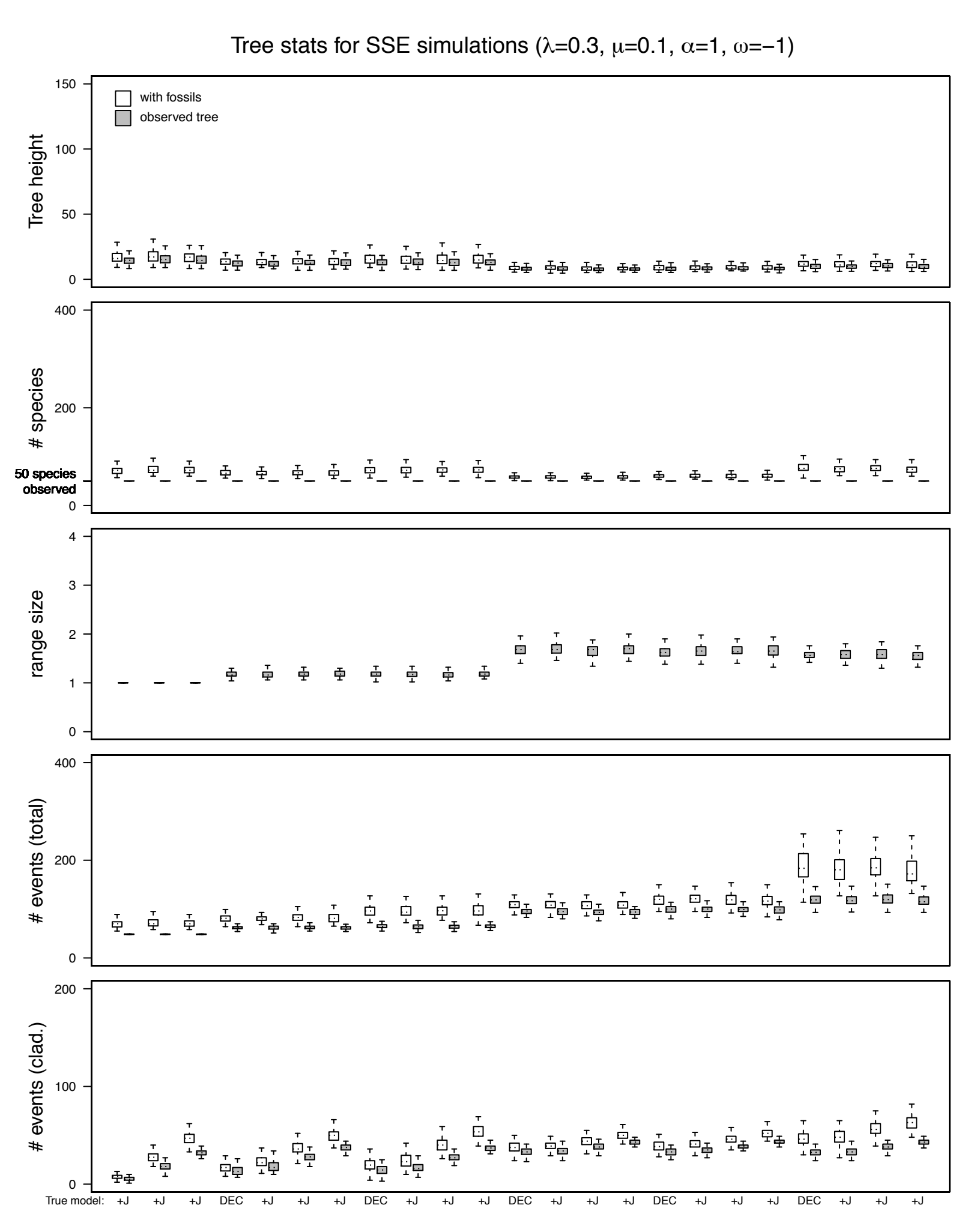
Supplemental Figure 8.2



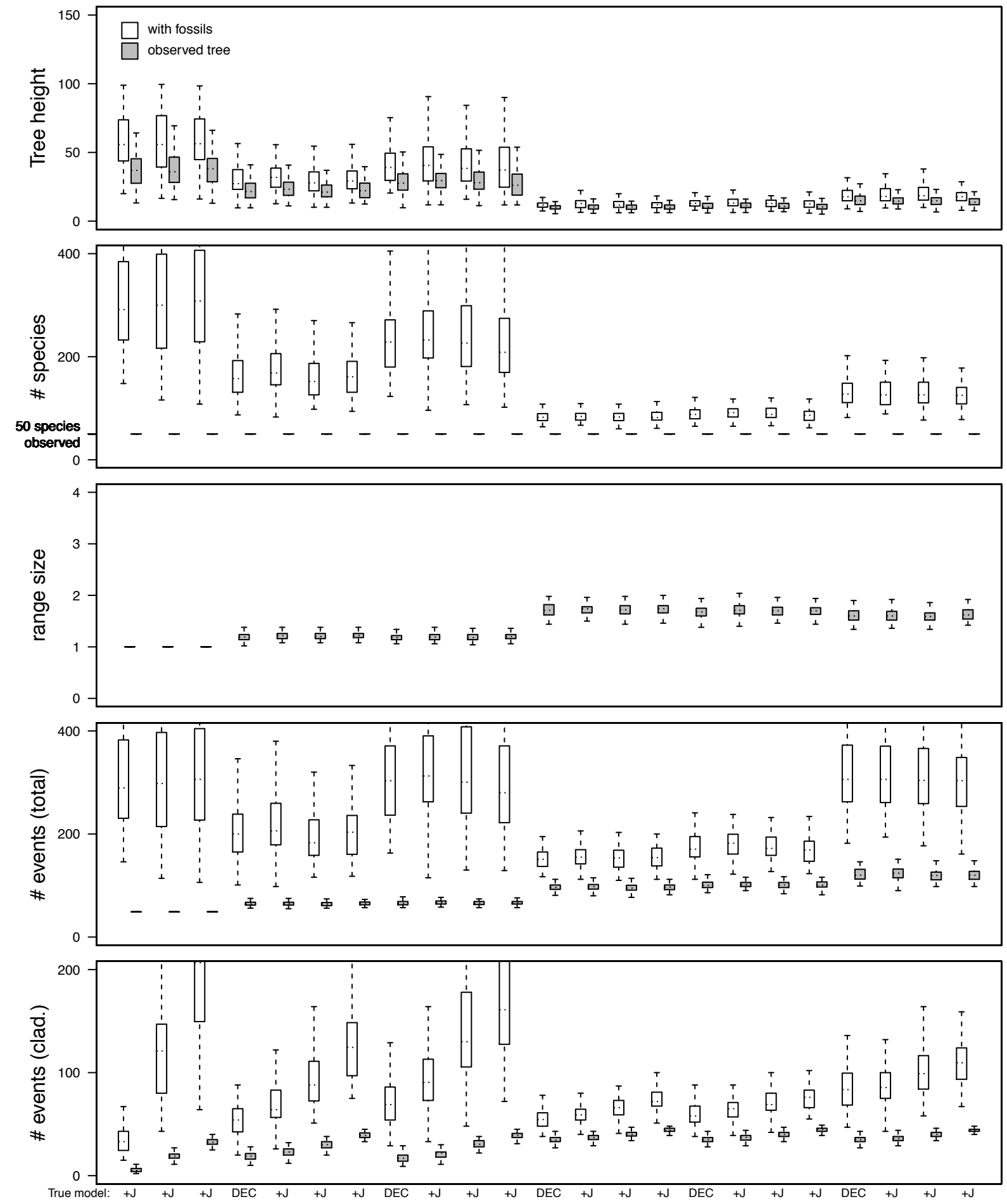
Supplemental Figure 8.3



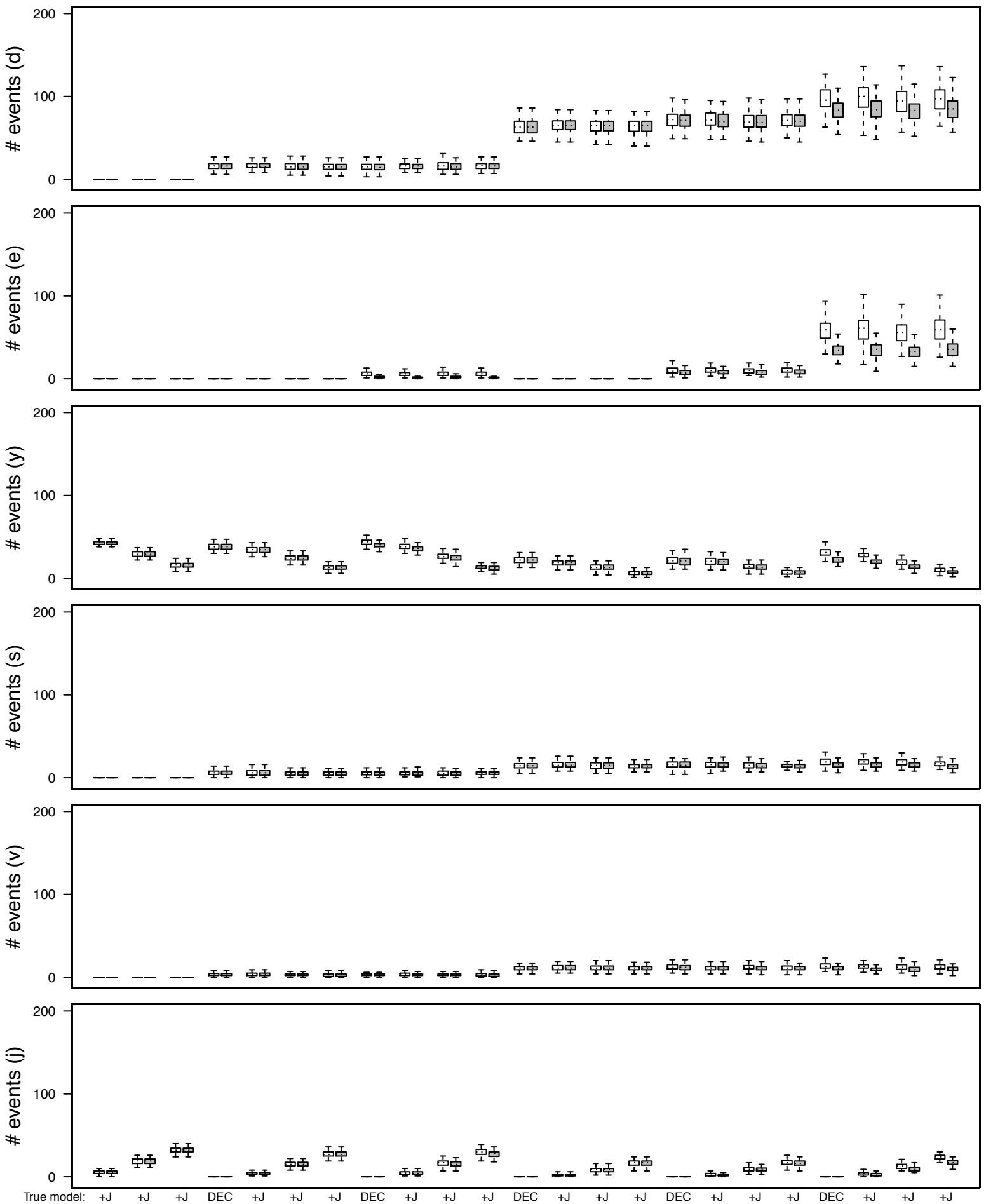
Supplemental Figure 8.4



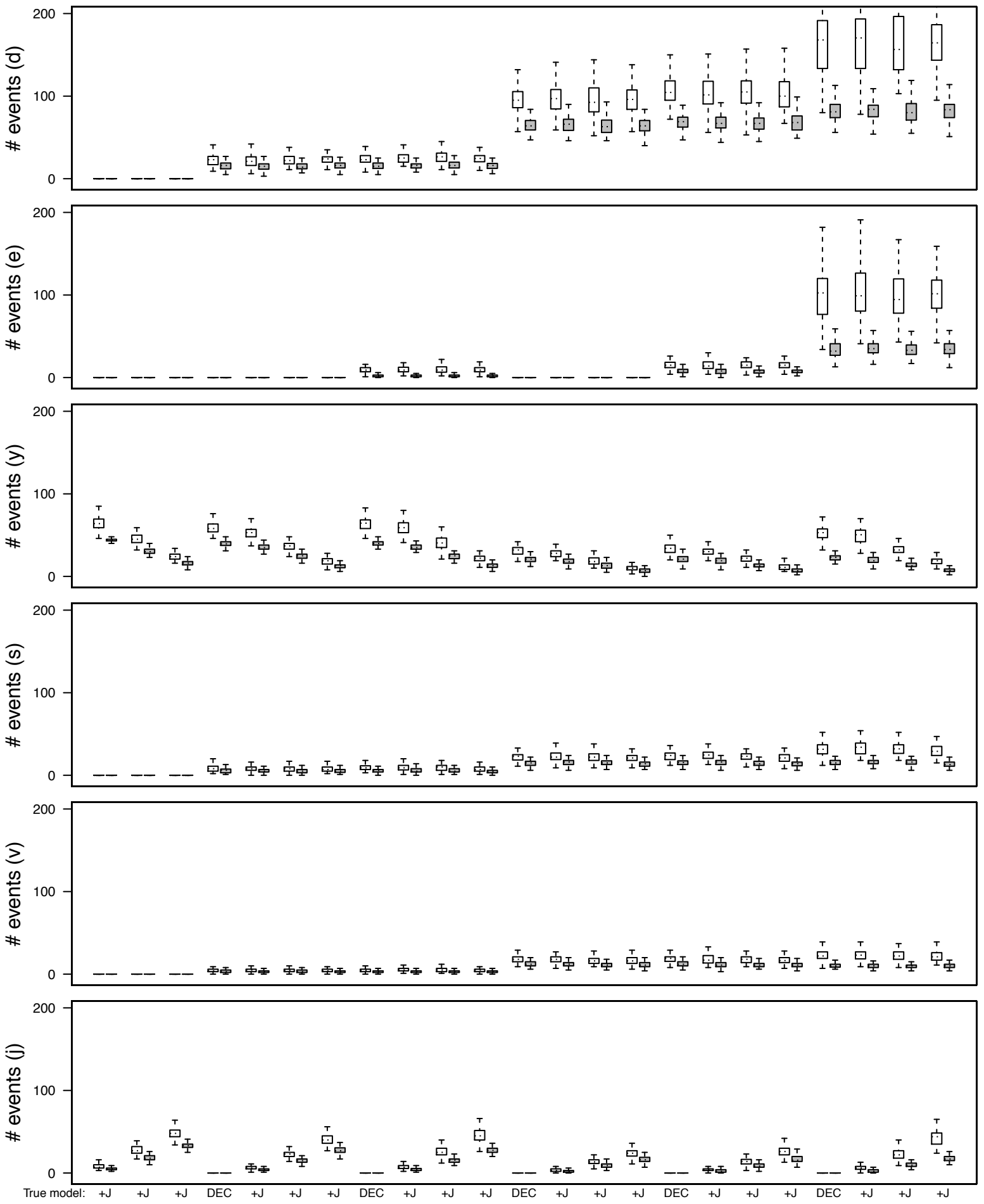
Supplemental Figure 8.5



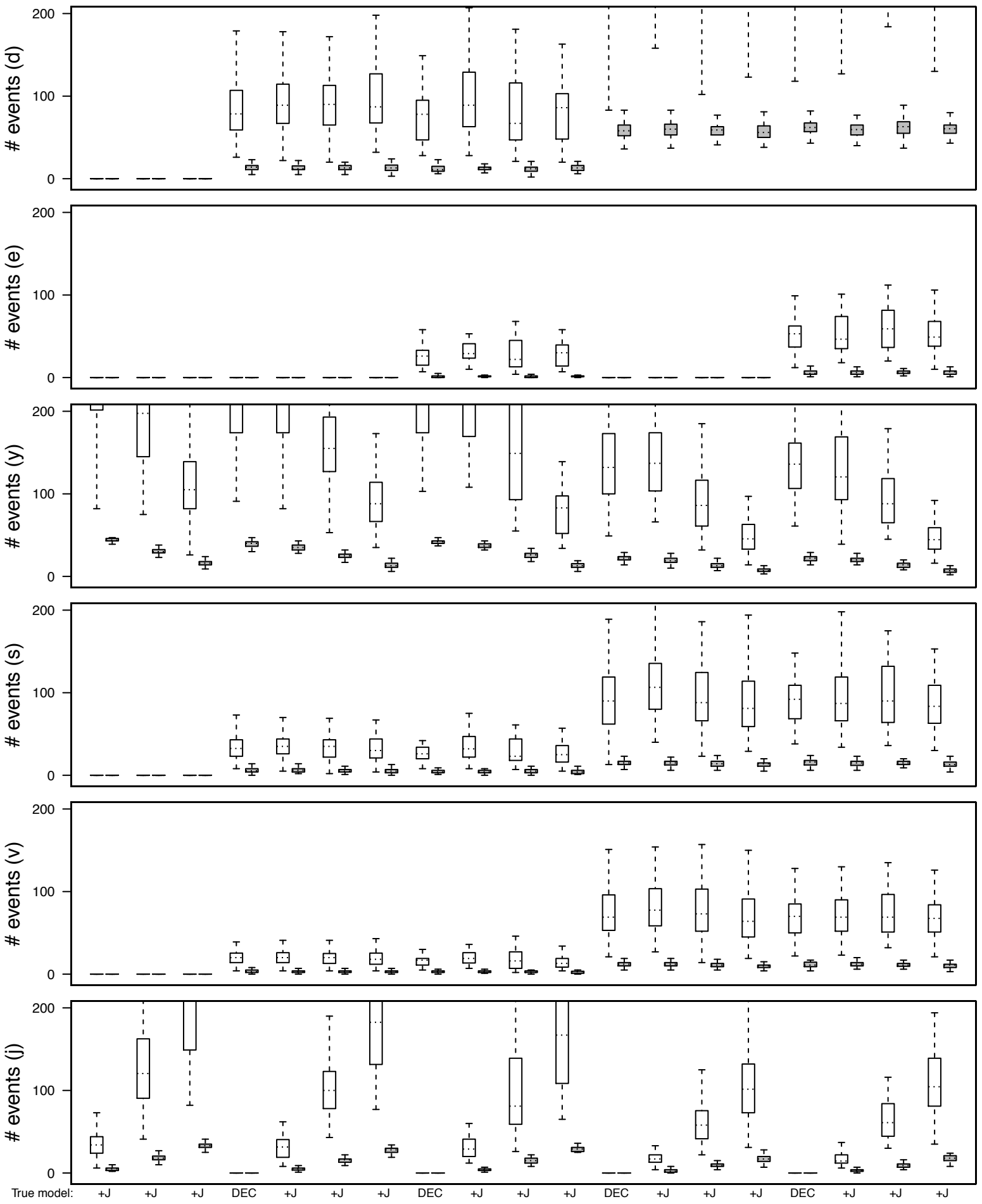
Supplemental Figure 8.6



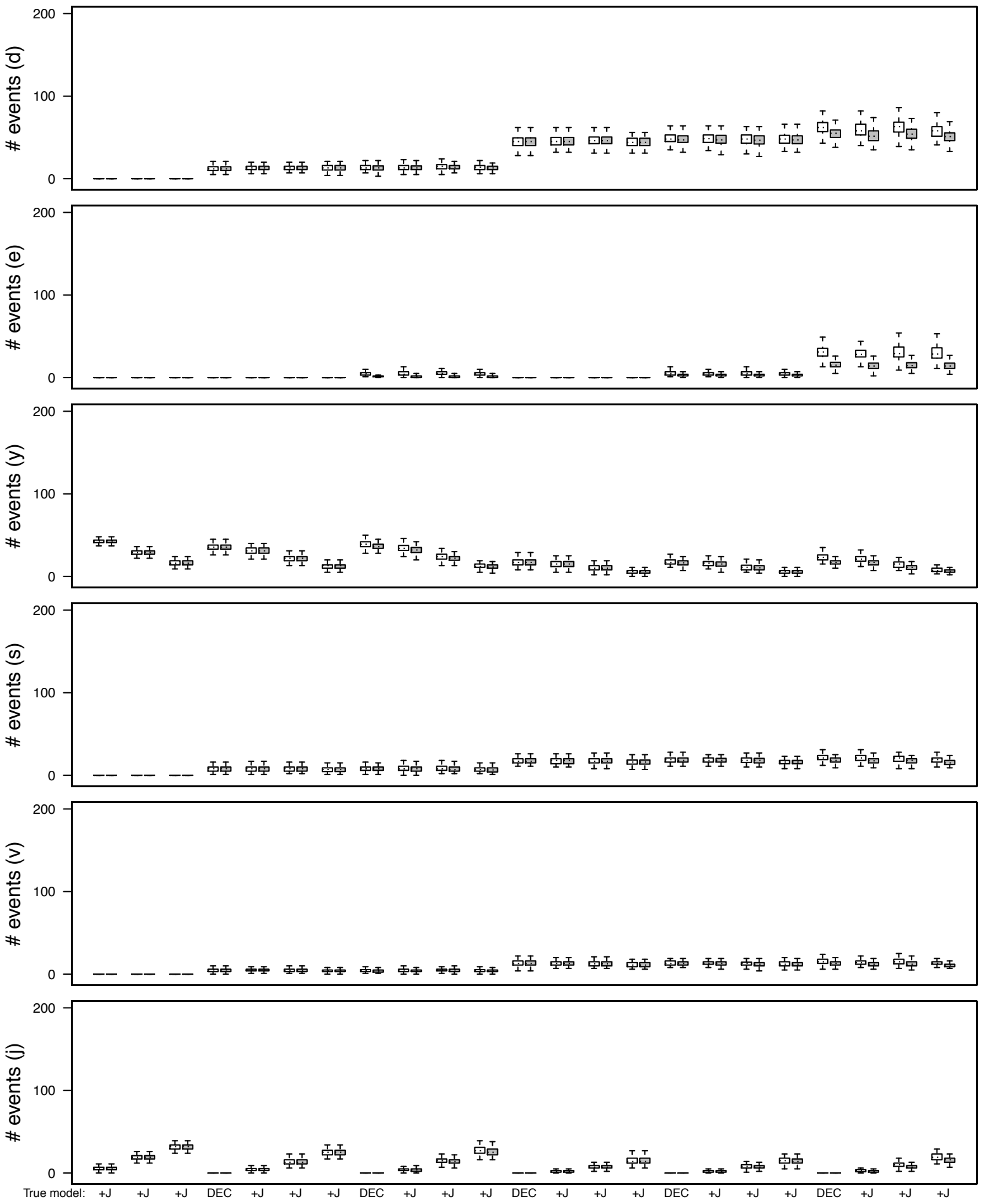
Supplemental Figure 8.7



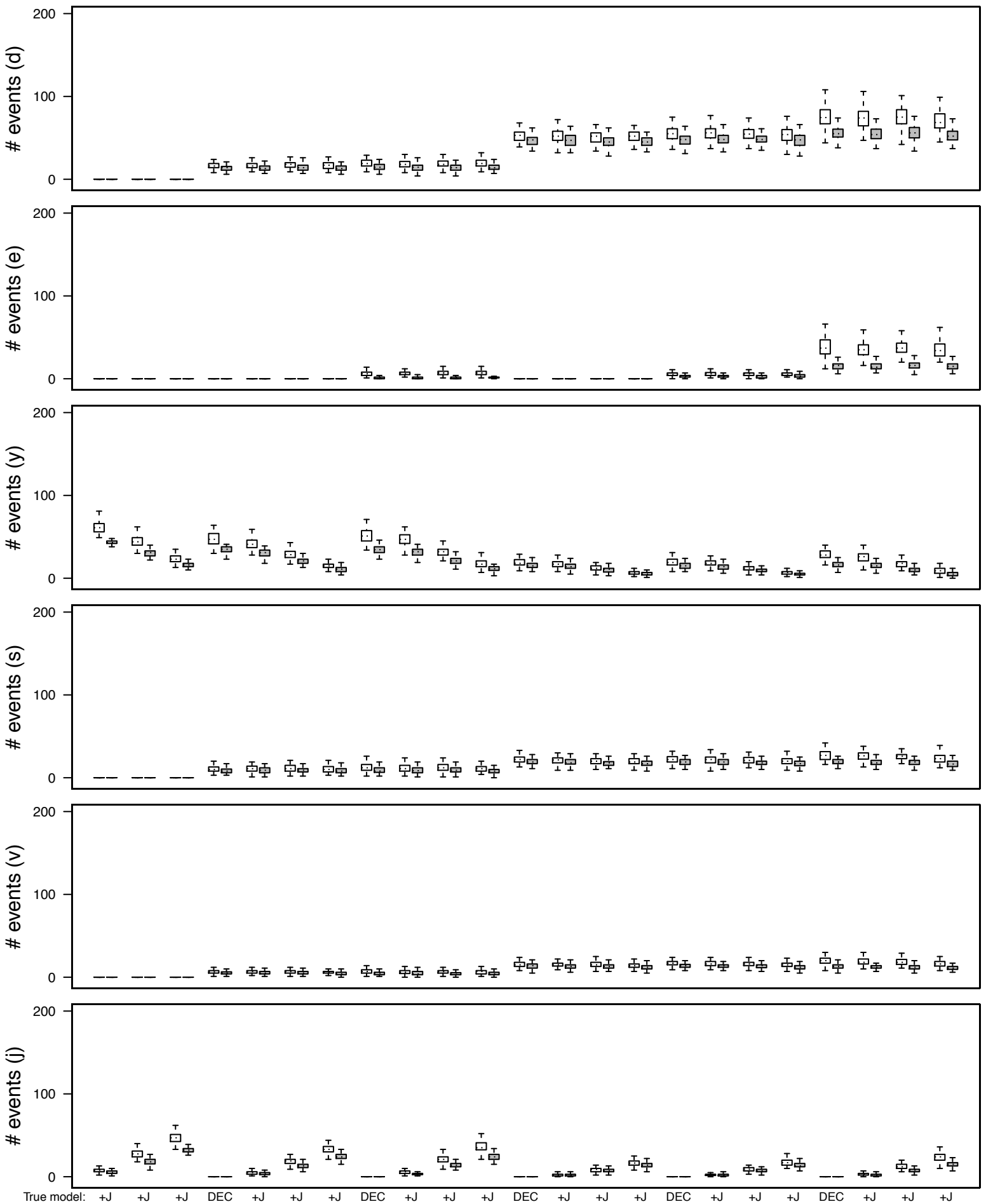
Supplemental Figure 8.8



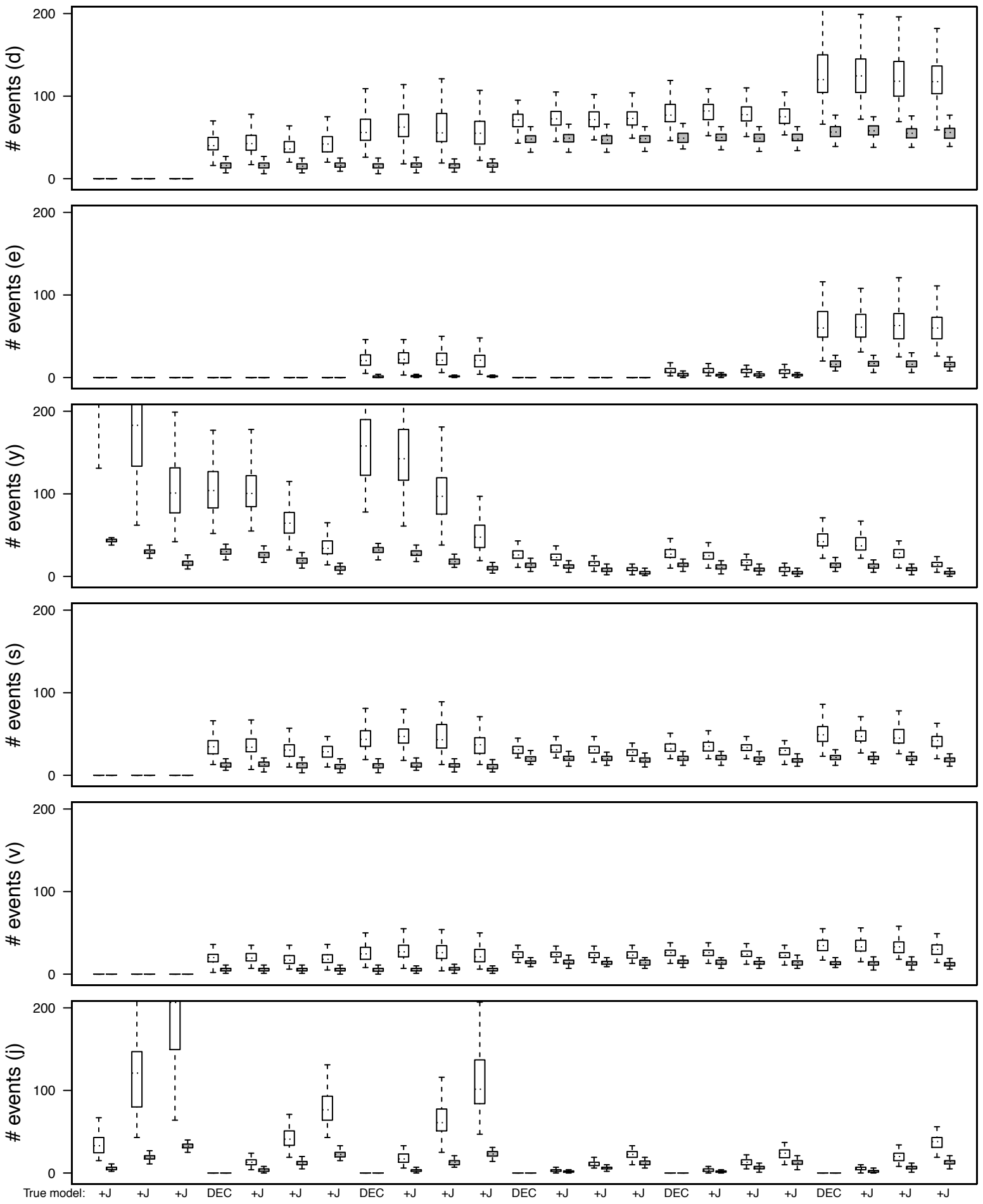
Supplemental Figure 8.9



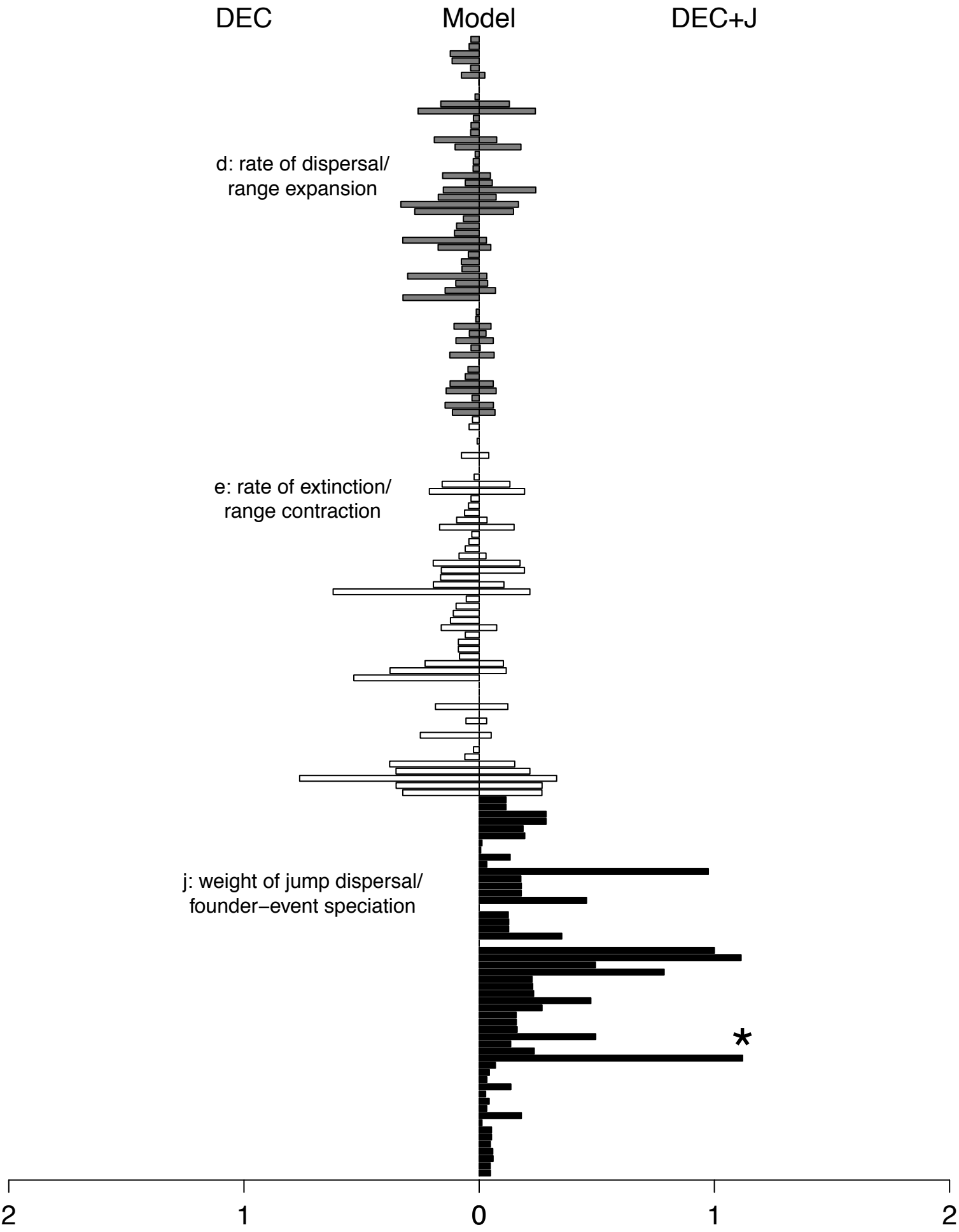
Supplemental Figure 8.10



Supplemental Figure 8.11



Supplemental Figure 8.12



Supplemental Figure 9