Statistical Topology and the Random Interstellar Medium

Supplementary Material

Contents

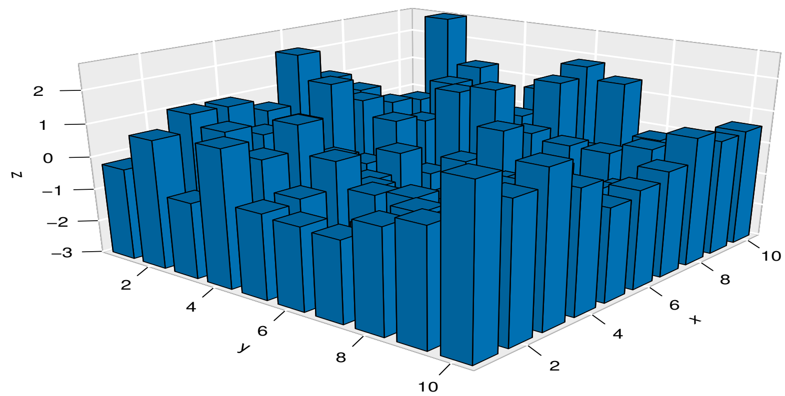
1 Additional plots

Figure 1 illustrates the data used in Section 1 of the main paper to describe level sets and Betti numbers.

Figures 2 and 3 show persistence diagrams and 90% convex peels for the three GASS data regions, after trend removal and marginal transformation.

Figure 4 provides an example simulation for each of the 5 models of Section 4.1.

(a)



(b)

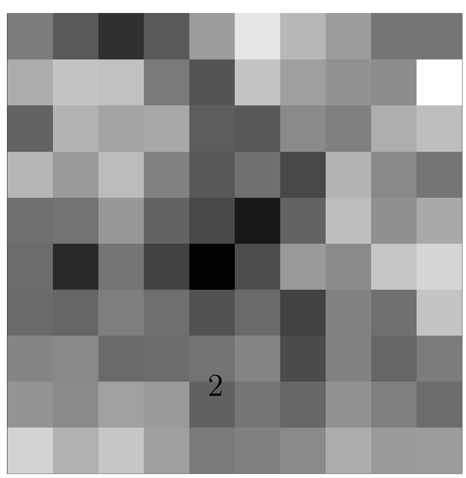


Figure 1: Two representations of the illustrative data used in Section 2.1 of the main paper

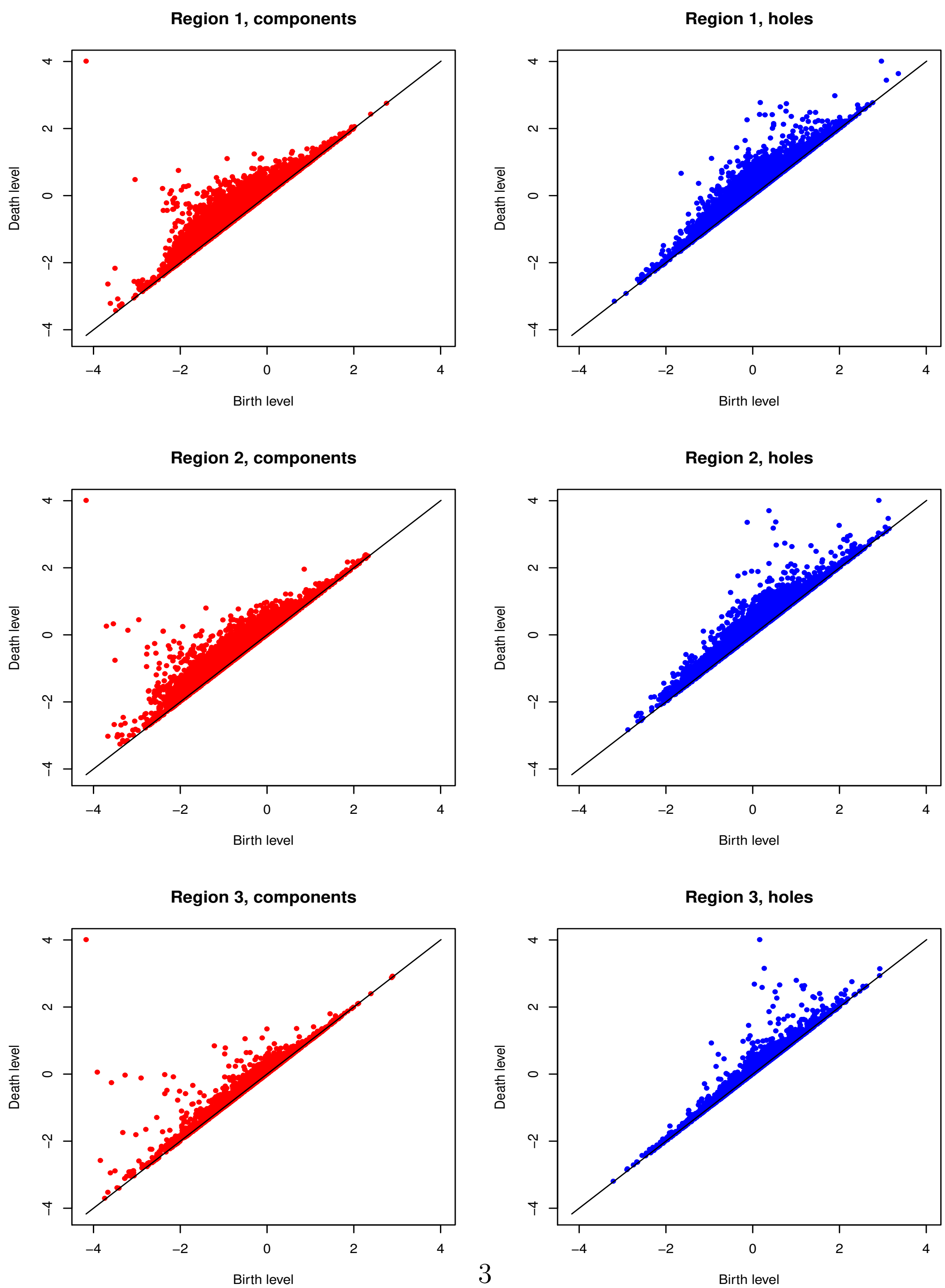


Figure 2: Persistence diagrams for GASS data regions

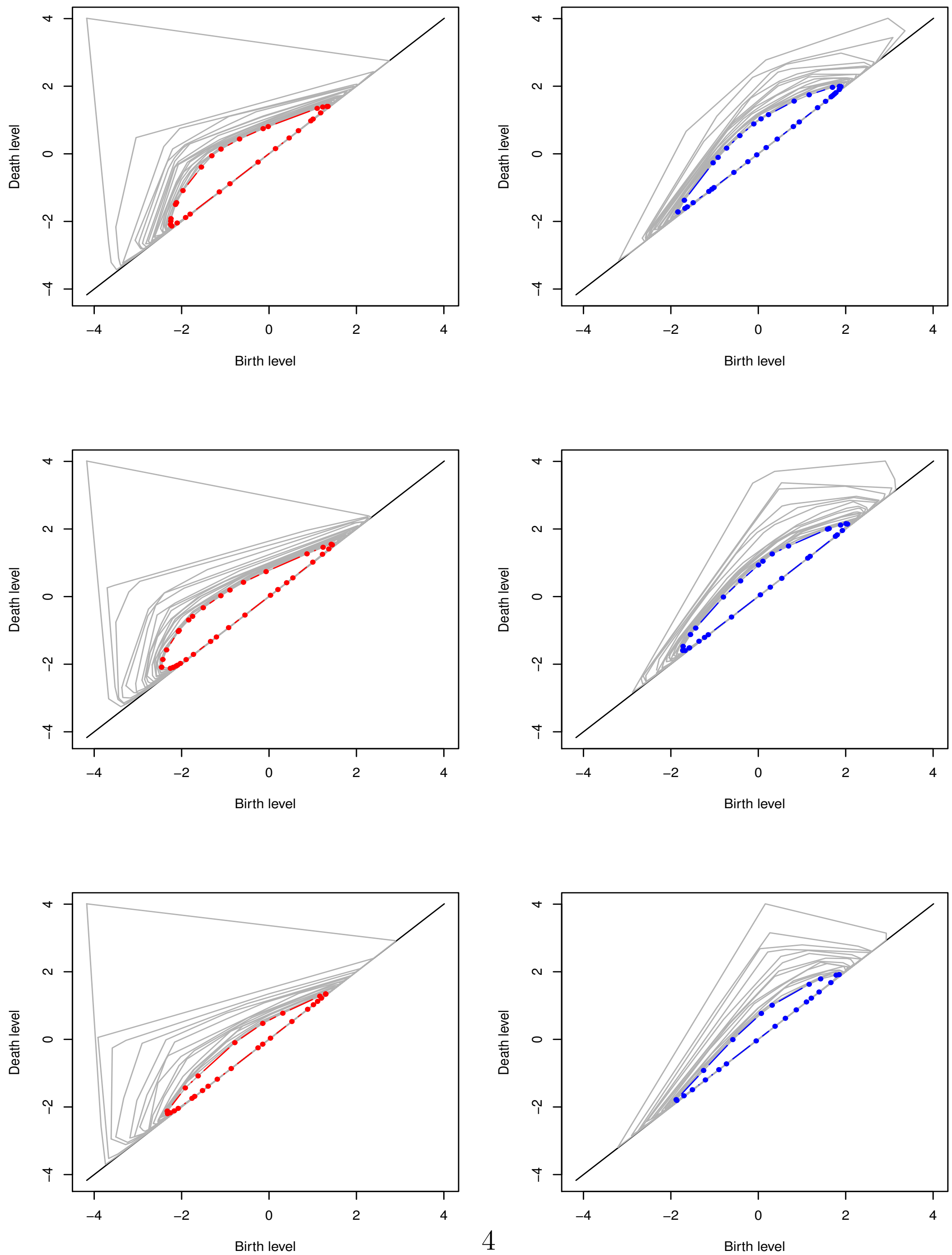


Figure 3: Convex peels for GASS data regions

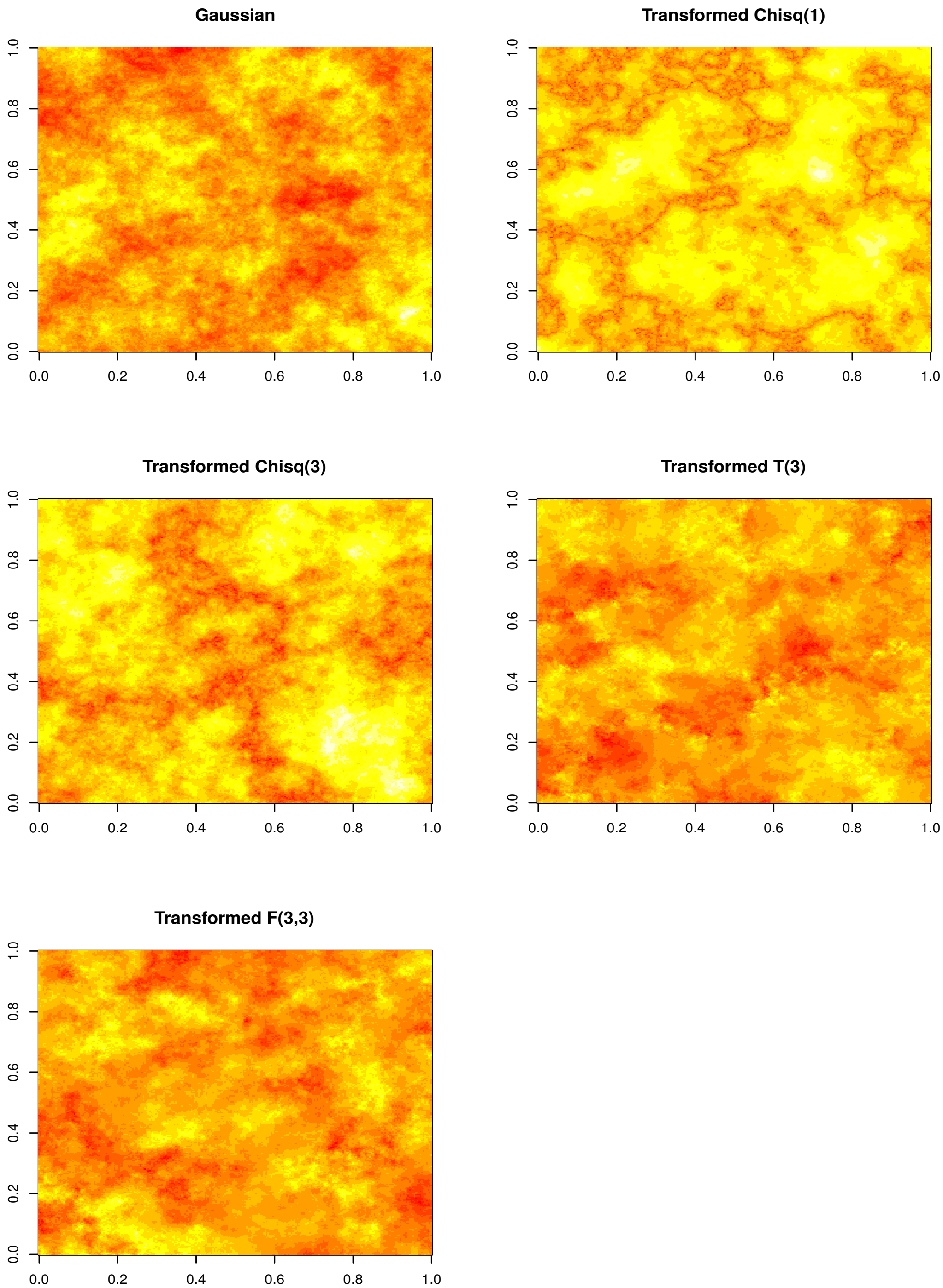


Figure 4: Examples of simulated data

2 Further simulation results

In Section 5 of the main paper we compare topologies of five different models, with (by construction) identical N(0,1) marginals and with correlation functions that are indistinguishable in practice. The first model is a Gaussian random field (GRF) with specified correlation function. The others are derived from GRFs but the constructed fields have intractable correlation functions. For comparability, we numerically searched for Matern correlation functions for the *root* GRFs that led to correlations in the *constructed* fields as close as possible to that for Model 1, with more weight on local correlations. The simulations in the paper assume exponential correlation for Model 1, based on that observed in ISM Regions 1 and 2. We repeat the construction and simulations here under the assumption that Model 1 has Matern correlation function based on that of Region 3.

Table 1: *Correlation in constructed fields compared with target correlation for Model 1: as Table 3 of main paper but with Matern correlation for Model 1, with parameters* $\nu = 1$ *and* $\eta = 15$*. The Matern parameters for the Gaussian random fields underpinning the other models were estimated numerically to match the correlations in the constructed fields to Model 1. For Models 2-5 the table shows mean (and standard deviation) of estimated correlations at various lags from 50 simulations of 256×256 random fields. For Model 1 the table shows the exact correlation.*

			Distance						
Model	ν	η	1	2	3	5	10	25	50
1: Gauss	1.00	15	0.987	0.959	0.924	0.841	0.626	0.208	0.026
2: χ^2_1	2.32	25	0.982	0.954	0.920	0.841	0.623	0.153	-0.008
			(0.003)	(0.007)	(0.012)	(0.022)	(0.049)	(0.081)	(0.068)
3: χ^2_3	0.97	27	0.986	0.956	0.919	0.832	0.607	0.179	-0.003
			(0.002)	(0.006)	(0.010)	(0.021)	(0.046)	(0.073)	(0.063)
4: T_3	1.32	16	0.988	0.961	0.924	0.837	0.613	0.182	-0.012
			(0.001)	(0.003)	(0.006)	(0.012)	(0.027)	(0.044	(0.065)
5: $F_{3,3}$	1.03	29	0.986	0.955	0.915	0.824	0.594	0.175	0.001
			(0.002)	(0.005)	(0.009)	(0.017)	(0.035)	(0.056)	(0.053)

With these root GRFs and constructed correlations, the nonparametric topological test sizes and power are estimated in the next table. Compared with Table 4 of the main paper, Model 3 is closer to Model 1 but further from Models 4 and 5. Test sizes are all good and in the main power is high.

Table 2: *As Table 4 of main paper but with correlation parameters as obtained above. The table gives test sizes and power estimates for nonparametric comparison between pairs of* 256×256 *random fields. Each value is obtained from 500 simulated pairs, using 5% overall tests after Bonferonni adjustment.*

	Gauss	χ^2_1	χ^2_3	T_3	$F_{3,3}$
Gauss	0.042	1.000	0.188	1.000	0.970
χ^2_1		0.060	1.000	1.000	1.000
χ^2_3			0.048	1.000	0.788
T_3				0.048	1.000
$F_{3,3}$					0.048

3 Comparison of observed component and hole counts with GRF expectations under different trend and covariance methods

The following tables compare the observed component and hole counts for the interstellar regions with expected numbers for Gaussian random fields with various correlation structures. We considered five different methods of removing trend from the interstellar data prior to topological analysis:

1. No trend removal at all.
2. A 4th order polynomial approach, using all terms of the form $b^i l^j$ with $i + j \leq 4$. Here b is galactic latitude and l is galactic longitude.
3. A higher order polynomial, as above but with $i + j \leq 7$.
4. A thin plate smoothing spline (TPSS) fit, with automatic smoothing parameter selection, as implemented in the `mgcv` R package.
5. As above but with higher order penalty (10 instead of the default 2).

We obtained the GRF expectations under five different correlation structures:

1. Fitting a Matern model to empirical correlations up to a separation of 70 grid units (long range Matern).
2. Fitting a Matern only to the local correlations, to separation 3 grid units (short range Matern).
3. Using empirically estimated local correlations, assuming stationarity and isotropy (Empirical: iso).
4. Using empirically estimated local correlations, assuming stationarity but not isotropy (Empirical: aniso).

5. Using empirically estimated local correlations, without assuming stationarity or isotropy (Empirical: nonstat).

For the non-stationarity results, we subdivided each field into 16 sub-regions and estimated the correlations empirically within each of these.

Results are given below for each region separately. If empirical correlations are used then the results are relatively insensitive to method of removing trend.

Region 1

Observed and expected counts

	No trend	Low poly	High poly	TPSS	TPSS-flexible
Observed components	5175	5201	5185	5189	5172
Observed holes	5081	5100	5145	5153	5125
Long range Matern	8841.8	7350.3	6305.1	6282.2	5890.1
Short range Matern	7663.9	6642.9	6715.0	6815.4	6677.7
Empirical: iso	5367.5	4965.3	5077.4	5121.0	5076.8
Empirical: aniso	5371.7	4965.6	5079.5	5122.3	5077.6
Emirical: non-stat	5215.4	5072.6	5157.9	5189.5	5131.7

Standard deviations

	No trend	Low poly	High poly	TPSS	TPSS-flexible
Long range Matern	66.2	65.6	64.2	64.2	63.3
Short range Matern	66.1	64.9	64.7	64.9	64.5
Empirical: iso	48.4	49.9	49.8	49.7	49.9
Empirical: aniso	50.3	50.0	49.9	50.0	49.9
Emirical: non-stat	51.5	51.2	51.3	51.3	51.3

Region 2

Observed and expected counts

	No trend	Low poly	High poly	TPSS	TPSS-flexible
Observed components	5092	5103	5086	5124	5070
Observed holes	5050	5095	5091	5115	5054
Long range Matern	9209.5	5738.6	3665.6	3248.6	2825.6
Short range Matern	7132.7	6968.6	6578.5	6612.1	6501.8
Empirical: iso	5358.0	5341.8	5132.4	5146.8	5123.6
Empirical: aniso	5358.2	5339.4	5130.4	5144.3	5121.4
Emirical: non-stat	5215.3	5160.5	5007.7	5065.5	4963.2

Standard deviations

	No trend	Low poly	High poly	TPSS	TPSS-flexible
Long range Matern	66.0	63.7	56.4	54.2	51.3
Short range Matern	65.7	65.4	64.8	64.8	64.5
Empirical: iso	50.6	50.9	51.0	51.0	51.1
Empirical: aniso	48.1	48.4	48.9	48.9	48.9
Emirical: non-stat	49.9	50.0	50.8	50.2	50.7

Region 3

Observed and expected counts

	No trend	Low poly	High poly	TPSS	TPSS-flexible
Observed components	3840	3831	3676	3831	3725
Observed holes	3694	3661	3519	3700	3565
Long range Matern	6709.4	2570.9	1833.3	1229.6	930.3
Short range Matern	6719.8	3411.8	2964.5	3428.2	2987.8
Empirical: iso	4883.7	2835.2	2603.9	2965.4	2650.2
Empirical: aniso	4879.7	2843.8	2613.1	2976.4	2662.6
Emirical: non-stat	4069.4	3125.8	2933.3	3295.2	2976.3

Standard deviations

	No trend	Low poly	High poly	TPSS	TPSS-flexible
Long range Matern	65.5	50.6	43.5	36.5	31.9
Short range Matern	65.5	55.9	52.8	55.5	52.6
Empirical: iso	49.9	48.3	47.2	49.1	47.5
Empirical: aniso	50.5	48.7	47.4	49.1	47.6
Emirical: non-stat	50.1	47.9	47.8	48.9	47.9

4 P-values for comparisons between regions under different trend models

The following tables are as Table 6 of the main paper, but using each of the five trend removal methods summarised above. These results do not rely on any assumptions about the correlation structure.

We see no evidence of differences between Region 1 and Region 2, but strong evidence that Region 3 differs from the other regions in numbers of components and in both numbers and filamentarity for holes.

Region 1 v Region 2

	Comps-filam	Comps-number	Holes-filam	Holes-number
No trend	0.730	0.860	0.666	1.000
Low poly	0.387	1.000	0.863	0.863
High poly	0.605	1.000	0.931	0.860
TPSS	0.931	0.860	0.666	0.730
TPSS-flexible	0.863	0.931	0.931	0.895

Region 1 v Region 3

	Comps-filam	Comps-number	Holes-filam	Holes-number
No trend	0.796	0.017	0.666	0.050
Low poly	0.077	0.019	0.019	0.027
High poly	0.863	0.008	0.014	0.013
TPSS	0.796	0.027	0.094	0.040
TPSS-flexible	0.387	0.009	0.050	0.012

Region 2 v Region 3

	Comps-filam	Comps-number	Holes-filam	Holes-number
No trend	0.931	0.031	0.605	0.063
Low poly	0.222	0.024	0.031	0.017
High poly	0.863	0.006	0.136	0.006
TPSS	0.796	0.019	0.340	0.030
TPSS-flexible	0.796	0.006	0.050	0.007