Peter Bhlmann

List of Publications by Citations

155 papers	15,993 citations	48 h-index	126 g-index
168 ext. papers	19,471 ext. citations	3.3 avg, IF	7.26 L-index

#	Paper	IF	Citations
155	High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics*, **2006**, 34, 1436	3.2	1622
154	MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics*, **2012**, 28, 112-8	7.2	1585
153	Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2010**, 72, 417-473	3.9	1221
152	Statistics for High-Dimensional Data. *Springer Series in Statistics*, **2011**,	0.3	853
151	The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2008**, 70, 53-71	3.9	813
150	A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, **2006**, 22, 1122-9	7.2	626
149	Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, **2007**, 23, 980-7	7.2	573
148	Boosting With the L2 Loss. *Journal of the American Statistical Association*, **2003**, 98, 324-339	2.8	485
147	Boosting Algorithms: Regularization, Prediction and Model Fitting. *Statistical Science*, **2007**, 22, 477	2.4	484
146	Survival ensembles. *Biostatistics*, **2006**, 7, 355-73	3.7	403
145	On asymptotically optimal confidence regions and tests for high-dimensional models. *Annals of Statistics*, **2014**, 42,	3.2	385
144	Analyzing bagging. *Annals of Statistics*, **2002**, 30, 927	3.2	383
143	On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, **2009**, 3,	1.2	290
142	Boosting for tumor classification with gene expression data. *Bioinformatics*, **2003**, 19, 1061-9	7.2	282
141	Sieve Bootstrap for Time Series. *Bernoulli*, **1997**, 3, 123	1.6	255
140	Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana. *Genome Biology*, **2004**, 5, R92	18.3	229
139	High-dimensional additive modeling. *Annals of Statistics*, **2009**, 37,	3.2	219

138	Boosting for high-dimensional linear models. *Annals of Statistics*, **2006**, 34, 559	3.2	216
137	p-Values for High-Dimensional Regression. *Journal of the American Statistical Association*, **2009**, 104, 1671-1681	2.8	205
136	Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. *Cancer Research*, **2004**, 64, 5539-45	10.1	196
135	Variable length Markov chains. *Annals of Statistics*, **1999**, 27, 480	3.2	196
134	Causal Inference Using Graphical Models with theRPackagepcalg. *Journal of Statistical Software*, **2012**, 47,	7.3	185
133	Targeted quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring. *Molecular and Cellular Proteomics*, **2008**, 7, 1489-500	7.6	179
132	Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. *Molecular Systems Biology*, **2012**, 8, 606	12.2	163
131	Bootstraps for Time Series. *Statistical Science*, **2002**, 17, 52	2.4	161
130	1-penalization for mixture regression models. *Test*, **2010**, 19, 209-256	1.1	143
129	Estimating high-dimensional intervention effects from observational data. *Annals of Statistics*, **2009**, 37,	3.2	142
128	Predicting causal effects in large-scale systems from observational data. *Nature Methods*, **2010**, 7, 247-8	21.6	136
127	Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2016**, 78, 947-1012	3.9	125
126	Selection of Carbonic Anhydrase IX Inhibitors from One Million DNA-Encoded Compounds. *ACS Chemical Biology*, **2011**, 6, 336-44	4.9	117
125	High-Dimensional Statistics with a View Toward Applications in Biology. *Annual Review of Statistics and Its Application*, **2014**, 1, 255-278	7.6	115
124	Statistical significance in high-dimensional linear models. *Bernoulli*, **2013**, 19,	1.6	106
123	Estimation for High-Dimensional Linear Mixed-Effects Models Using 1-Penalization. *Scandinavian Journal of Statistics*, **2011**, 38, 197-214	0.8	97
122	Discovery of TNF inhibitors from a DNA-encoded chemical library based on diels-alder cycloaddition. *Chemistry and Biology*, **2009**, 16, 1075-86		89
121	Correlated variables in regression: Clustering and sparse estimation. *Journal of Statistical Planning and Inference*, **2013**, 143, 1835-1858	0.8	86

120	High-Dimensional Inference: Confidence Intervals, p-Values and R-Software hdi. *Statistical Science*, **2015**, 30,	2.4	84
119	Bagging, Boosting and Ensemble Methods **2012**, 985-1022		83
118	Finding predictive gene groups from microarray data. *Journal of Multivariate Analysis*, **2004**, 90, 106-131	1.4	80
117	Block length selection in the bootstrap for time series. *Computational Statistics and Data Analysis*, **1999**, 31, 295-310	1.6	77
116	Arabidopsis GERANYLGERANYL DIPHOSPHATE SYNTHASE 11 is a hub isozyme required for the production of most photosynthesis-related isoprenoids. *New Phytologist*, **2016**, 209, 252-64	9.8	73
115	Mining tissue microarray data to uncover combinations of biomarker expression patterns that improve intermediate staging and grading of clear cell renal cell cancer. *Clinical Cancer Research*, **2010**, 16, 88-98	12.9	70
114	Low-order conditional independence graphs for inferring genetic networks. *Statistical Applications in Genetics and Molecular Biology*, **2006**, 5, Article1	1.2	63
113	CAM: Causal additive models, high-dimensional order search and penalized regression. *Annals of Statistics*, **2014**, 42,	3.2	62
112	Identifiability of Gaussian structural equation models with equal error variances. *Biometrika*, **2014**, 101, 219-228	2	62
111	Blockwise Bootstrapped Empirical Process for Stationary Sequences. *Annals of Statistics*, **1994**, 22, 995	3.2	59
110	The adaptive and the thresholded Lasso for potentially misspecified models (and a lower bound for the Lasso). *Electronic Journal of Statistics*, **2011**, 5,	1.2	52
109	Kernel-based tests for joint independence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2018**, 80, 5-31	3.9	50
108	Variable selection in high-dimensional linear models: partially faithful distributions and the PC-simple algorithm. *Biometrika*, **2010**, 97, 261-278	2	49
107	GLMMLasso: An Algorithm for High-Dimensional Generalized Linear Mixed Models Using ℓ_1-Penalization. *Journal of Computational and Graphical Statistics*, **2014**, 23, 460-477	1.4	48
106	Geometry of the faithfulness assumption in causal inference. *Annals of Statistics*, **2013**, 41,	3.2	48
105	Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences of the United States of America*, **2016**, 113, 7361-8	11.5	48
104	A New Mixing Notion and Functional Central Limit Theorems for a Sieve Bootstrap in Time Series. *Bernoulli*, **1999**, 5, 413	1.6	47
103	Sieve bootstrap for smoothing in nonstationary time series. *Annals of Statistics*, **1998**, 26, 48	3.2	46

102	An algorithm for nonparametric GARCH modelling. *Computational Statistics and Data Analysis*, **2002**, 40, 665-683	1.6	44
101	ell_{0}-penalized maximum likelihood for sparse directed acyclic graphs. *Annals of Statistics*, **2013**, 41,	3.2	43
100	Conditional transformation models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2014**, 76, 3-27	3.9	42
99	Twin Boosting: improved feature selection and prediction. *Statistics and Computing*, **2010**, 20, 119-138	1.8	40
98	Causal stability ranking. *Bioinformatics*, **2012**, 28, 2819-23	7.2	38
97	Model-based boosting in high dimensions. *Bioinformatics*, **2006**, 22, 2828-9	7.2	38
96	Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics*, **2002**, 30, 397	3.2	38
95	High-dimensional simultaneous inference with the bootstrap. *Test*, **2017**, 26, 685-719	1.1	37
94	Most Likely Transformations. *Scandinavian Journal of Statistics*, **2018**, 45, 110-134	0.8	37
93	Stable graphical model estimation with Random Forests for discrete, continuous, and mixed variables. *Computational Statistics and Data Analysis*, **2013**, 64, 132-152	1.6	37
92	LOCALLY ADAPTIVE LAG-WINDOW SPECTRAL ESTIMATION. *Journal of Time Series Analysis*, **1996**, 17, 247-270	0.8	37
91	High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Computational Statistics*, **2014**, 29, 407-430	1	35
90	Variable Length Markov Chains: Methodology, Computing, and Software. *Journal of Computational and Graphical Statistics*, **2004**, 13, 435-455	1.4	34
89	Protein and gene model inference based on statistical modeling in k-partite graphs. *Proceedings of the National Academy of Sciences of the United States of America*, **2010**, 107, 12101-6	11.5	33
88	Asymptotic optimality of the Westfall Young permutation procedure for multiple testing under dependence. *Annals of Statistics*, **2011**, 39,	3.2	32
87	Understanding human functioning using graphical models. *BMC Medical Research Methodology*, **2010**, 10, 14	4.7	31
86	Discussion: One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, **2008**, 36,	3.2	30
85	Tree-structured generalized autoregressive conditional heteroscedastic models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2001**, 63, 727-744	3.9	29

84	A multi-marker association method for genome-wide association studies without the need for population structure correction. *Nature Communications*, **2016**, 7, 13299	17.4	28
83	Simultaneous analysis of large-scale RNAi screens for pathogen entry. *BMC Genomics*, **2014**, 15, 1162	4.5	28
82	Missing values: sparse inverse covariance estimation and an extension to sparse regression. *Statistics and Computing*, **2012**, 22, 219-235	1.8	27
81	Robustification of the PC-Algorithm for Directed Acyclic Graphs. *Journal of Computational and Graphical Statistics*, **2008**, 17, 773-789	1.4	27
80	Jointly interventional and observational data: estimation of interventional Markov equivalence classes of directed acyclic graphs. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2015**, 77, 291-318	3.9	26
79	Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika*, **2005**, 92, 893-907	2	26
78	Moving-average representation of autoregressive approximations. *Stochastic Processes and Their Applications*, **1995**, 60, 331-342	1.1	26
77	Bagging, Subagging and Bragging for Improving some Prediction Algorithms **2003**, 19-34		25
76	Dynamic adaptive partitioning for nonlinear time series. *Biometrika*, **1999**, 86, 555-571	2	25
75	Volatility estimation with functional gradient descent for very high-dimensional financial time series. *Journal of Computational Finance*, **2003**, 6, 65-89	1.7	25
74	Statistical approach to protein quantification. *Molecular and Cellular Proteomics*, **2014**, 13, 666-77	7.6	23
73	Integrative genome-wide expression profiling identifies three distinct molecular subgroups of renal cell carcinoma with different patient outcome. *BMC Cancer*, **2012**, 12, 310	4.8	23
72	Robustified L2 boosting. *Computational Statistics and Data Analysis*, **2008**, 52, 3331-3341	1.6	23
71	Assessing statistical significance in multivariable genome wide association analysis. *Bioinformatics*, **2016**, 32, 1990-2000	7.2	22
70	Model Selection for Variable Length Markov Chains and Tuning the Context Algorithm. *Annals of the Institute of Statistical Mathematics*, **2000**, 52, 287-315	1	22
69	Goodness-of-fit tests for high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2018**, 80, 113-135	3.9	21
68	What is a linear process?. *Proceedings of the National Academy of Sciences of the United States of America*, **1996**, 93, 12128-31	11.5	21
67	Invariant Causal Prediction for Sequential Data. *Journal of the American Statistical Association*, **2019** , 114, 1264-1276	2.8	20

66	Maximin effects in inhomogeneous large-scale data. *Annals of Statistics*, **2015**, 43,	3.2	20
65	High-dimensional inference in misspecified linear models. *Electronic Journal of Statistics*, **2015**, 9,	1.2	20
64	The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and Their Applications*, **1995**, 58, 247-265	1.1	19
63	Closure of Linear Processes. *Journal of Theoretical Probability*, **1997**, 10, 445-479	0.5	18
62	Statistics for big data: A perspective. *Statistics and Probability Letters*, **2018**, 136, 37-41	0.6	17
61	Magging: Maximin Aggregation for Inhomogeneous Large-Scale Data. *Proceedings of the IEEE*, **2016**, 104, 126-135	14.3	17
60	Penalized likelihood for sparse contingency tables with an application to full-length cDNA libraries. *BMC Bioinformatics*, **2007**, 8, 476	3.6	16
59	Discussion of "the evolution of boosting algorithms" and "extending statistical boosting". *Methods of Information in Medicine*, **2014**, 53, 436-45	1.5	15
58	Splines for financial volatility. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2009**, 71, 655-670	3.9	14
57	Smoothing 1-penalized estimators for high-dimensional time-course data. *Electronic Journal of Statistics*, **2007**, 1,	1.2	14
56	High dimensional sparse covariance estimation via directed acyclic graphs. *Electronic Journal of Statistics*, **2009**, 3,	1.2	14
55	Causal statistical inference in high dimensions. *Mathematical Methods of Operations Research*, **2013**, 77, 357-370	1	13
54	Structural intervention distance for evaluating causal graphs. *Neural Computation*, **2015**, 27, 771-99	2.9	12
53	Hierarchical Testing in the High-Dimensional Setting With Correlated Variables. *Journal of the American Statistical Association*, **2016**, 111, 331-343	2.8	12
52	Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, **2014**, 55, 926-939	3.6	12
51	Sieve Bootstrap With Variable-Length Markov Chains for Stationary Categorical Time Series. *Journal of the American Statistical Association*, **2002**, 97, 443-471	2.8	12
50	Causal Structure Learning and Inference: A Selective Review. *Quality Technology and Quantitative Management*, **2014**, 11, 3-21	1.9	11
49	Extreme events from the return-volume process: a discretization approach for complexity reduction. *Applied Financial Economics*, **1998**, 8, 267-278		11

48	Invariance, Causality and Robustness. *Statistical Science*, **2020**, 35,	2.4	10
47	Anchor regression: Heterogeneous data meet causality. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2021**, 83, 215-246	3.9	10
46	Decomposition and model selection for large contingency tables. *Biometrical Journal*, **2010**, 52, 233-52	1.5	9
45	Score-based causal learning in additive noise models. *Statistics*, **2016**, 50, 471-485	0.5	9
44	Goodness-of-fit testing in high dimensional generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **2020**, 82, 773-795	3.9	8
43	Controlling false positive selections in high-dimensional regression and causal inference. *Statistical Methods in Medical Research*, **2013**, 22, 466-92	2.3	8
42	Boosting. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2010**, 2, 69-74	1.4	8
41	Efficient and adaptive post-model-selection estimators. *Journal of Statistical Planning and Inference* , **1999**, 79, 1-9	0.8	8
40	Causal Dantzig: Fast inference in linear structural equation models with hidden variables under additive interventions. *Annals of Statistics*, **2019**, 47,	3.2	7
39	Discussion: A significance test for the lasso *Annals of Statistics*, **2014**, 42,	3.2	6
38	Annotating novel genes by integrating synthetic lethals and genomic information. *BMC Systems Biology*, **2008**, 2, 3	3.5	6
37	Hierarchical inference for genome-wide association studies: a view on methodology with software. *Computational Statistics*, **2020**, 35, 1-40	1	5
36	Rejoinder: 1-penalization for mixture regression models. *Test*, **2010**, 19, 280-285	1.1	5
35	EVE (external variance estimation) increases statistical power for detecting differentially expressed genes. *Plant Journal*, **2007**, 52, 561-9	6.9	5
34	Conjugate Direction Boosting. *Journal of Computational and Graphical Statistics*, **2006**, 15, 287-311	1.4	5
33	Nonparametric causal inference from observational time series through marginal integration. *Econometrics and Statistics*, **2017**, 2, 81-105	0.8	4
32	A Sequential Rejection Testing Method for High-Dimensional Regression with Correlated Variables. *International Journal of Biostatistics*, **2016**, 12, 79-95	1.3	4
31	Marginal integration for nonparametric causal inference. *Electronic Journal of Statistics*, **2015**, 9,	1.2	4

30	Statistical Analysis of Quantum Chemical Data Using Generalized XML/CML Archives for the Derivation of Molecular Design Rules. *Chimia*, **2007**, 61, 165-168	1.3	4
29	Network analysis of systems elements. *Exs*, **2007**, 97, 331-51		4
28	A Look at Robustness and Stability of ell_{1}-versus ell_{0}-Regularization: Discussion of Papers by Bertsimas et al. and Hastie et al.. *Statistical Science*, **2020**, 35,	2.4	4
27	Hypersurfaces and Their Singularities in Partial Correlation Testing. *Foundations of Computational Mathematics*, **2014**, 14, 1079-1116	2.7	3
26	Multi-Omic Profiling of the Liver Across Diets and Age in a Diverse Mouse Population		3
25	Causal inference in partially linear structural equation models. *Annals of Statistics*, **2018**, 46,	3.2	3
24	High-dimensional statistics, with applications to genome-wide association studies. *EMS Surveys in Mathematical Sciences*, **2017**, 4, 45-75	1.4	2
23	Remembrance of Leo Breiman. *Annals of Applied Statistics*, **2010**, 4,	2.1	2
22	Prediction of Spatial Cumulative Distribution Functions Using Subsampling: Comment. *Journal of the American Statistical Association*, **1999**, 94, 97	2.8	2
21	Discussion of: Treelets—An adaptive multi-scale basis for sparse unordered data. *Annals of Applied Statistics*, **2008**, 2,	2.1	2
20	Multiomic profiling of the liver across diets and age in a diverse mouse population. *Cell Systems*, **2021**,	10.6	2
19	Deconfounding and Causal Regularisation for Stability and External Validity. *International Statistical Review*, **2020**, 88, S114	1.4	2
18	Change-Point Detection for Graphical Models in the Presence of Missing Values. *Journal of Computational and Graphical Statistics*, **2021**, 1-12	1.4	2
17	Rejoinder on: High-dimensional simultaneous inference with the bootstrap. *Test*, **2017**, 26, 751-758	1.1	1
16	Comments on: Data science, big data and statistics. *Test*, **2019**, 28, 330-333	1.1	1
15	Some Themes in High-Dimensional Statistics. *Abel Symposia*, **2016**, 1-13	0.9	1
14	Discussion of Big Bayes Stories and BayesBag. *Statistical Science*, **2014**, 29,	2.4	1
13	Rejoinder: Invariance, Causality and Robustness. *Statistical Science*, **2020**, 35,	2.4	1

12	Boosting Algorithms: with an Application to Bootstrapping Multivariate Time Series **2006**, 209-230		1
11	Robust Statistics **2014**, 51-98		1
10	Stabilizing variable selection and regression. *Annals of Applied Statistics*, **2021**, 15,	2.1	1
9	Seeded intervals and noise level estimation in change point detection: a discussion of Fryzlewicz (2020). *Journal of the Korean Statistical Society*, **2020**, 49, 1081-1089	0.5	0
8	Partial Least Squares for Heterogeneous Data. *Springer Proceedings in Mathematics and Statistics*, **2016**, 3-15	0.2	0
7	Toward causality and improving external validity. *Proceedings of the National Academy of Sciences of the United States of America*, **2020**, 117, 25963-25965	11.5	0
6	SPHN/PHRT: Forming a Swiss-Wide Infrastructure for Data-Driven Sepsis Research. *Studies in Health Technology and Informatics*, **2020**, 270, 1163-1167	0.5	0
5	Distributional anchor regression.. *Statistics and Computing*, **2022**, 32, 39	1.8	0
4	Rejoinder on: Hierarchical inference for genome-wide association studies: a view on methodology with software. *Computational Statistics*, **2020**, 35, 59-67	1	
3	Comments on: A random forest guided tour. *Test*, **2016**, 25, 239-246	1.1	
2	Boosting and ℓ_1-Penalty Methods for High-dimensional Data with Some Applications in Genomics **2006**, 1-12		
1	Confidence Intervals and Tests for High-Dimensional Models: A Compact Review. *Lecture Notes in Statistics*, **2015**, 21-34	2.9	