

# Bibliography

- Abrahamsen, P. (1997). A Review of Gaussian Random Fields and Correlation Functions. Technical Report 917, Norwegian Computing Center, Oslo, Norway. [http://publications.nr.no/917\\_Rapport.pdf](http://publications.nr.no/917_Rapport.pdf). p. 82
- Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions*. Dover, New York. pp. 84, 85
- Adams, R. (1975). *Sobolev Spaces*. Academic Press, New York. p. 134
- Adler, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester. pp. 80, 81, 83, 191, 218
- Amari, S. (1985). *Differential-Geometrical Methods in Statistics*. Springer-Verlag, Berlin. p. 102
- Ansley, C. F. and Kohn, R. (1985). Estimation, Filtering, and Smoothing in State Space Models with Incompletely Specified Initial Conditions. *Annals of Statistics*, 13(4):1286–1316. p. 29
- Arató, M. (1982). *Linear Stochastic Systems with Constant Coefficients*. Springer-Verlag, Berlin. Lecture Notes in Control and Information Sciences 45. p. 212
- Arfken, G. (1985). *Mathematical Methods for Physicists*. Academic Press, San Diego. pp. xv, 134
- Aronszajn, N. (1950). Theory of Reproducing Kernels. *Trans. Amer. Math. Soc.*, 68:337–404. pp. 129, 130
- Bach, F. R. and Jordan, M. I. (2002). Kernel Independent Component Analysis. *Journal of Machine Learning Research*, 3(1):1–48. p. 97
- Baker, C. T. H. (1977). *The Numerical Treatment of Integral Equations*. Clarendon Press, Oxford. pp. 98, 99
- Barber, D. and Saad, D. (1996). Does Extra Knowledge Necessarily Improve Generalisation? *Neural Computation*, 8:202–214. p. 31
- Bartle, R. G. (1995). *The Elements of Integration and Lebesgue Measure*. Wiley, New York. p. 204
- Bartlett, P. L., Jordan, M. I., and McAuliffe, J. D. (2003). Convexity, Classification and Risk Bounds. Technical Report 638, Department of Statistics, University of California, Berkeley. Available from <http://www.stat.berkeley.edu/tech-reports/638.pdf>. Accepted for publication in Journal of the American Statistical Association. p. 157
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York. Second edition. pp. 22, 35

- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford. p. 45
- Bishop, C. M., Svensen, M., and Williams, C. K. I. (1998a). Developments of the Generative Topographic Mapping. *Neurocomputing*, 21:203–224. p. 196
- Bishop, C. M., Svensen, M., and Williams, C. K. I. (1998b). GTM: The Generative Topographic Mapping. *Neural Computation*, 10(1):215–234. p. 196
- Blake, I. F. and Lindsey, W. C. (1973). Level-Crossing Problems for Random Processes. *IEEE Trans Information Theory*, 19(3):295–315. p. 81
- Blight, B. J. N. and Ott, L. (1975). A Bayesian Approach to Model Inadequacy for Polynomial Regression. *Biometrika*, 62(1):79–88. p. 28
- Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, Cambridge, UK. p. 206
- Boyle, P. and Frean, M. (2005). Dependent Gaussian Processes. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 217–224. MIT Press. p. 190
- Bracewell, R. N. (1986). *The Fourier Transform and Its Applications*. McGraw-Hill, Singapore, international edition. pp. 83, 206
- Caruana, R. (1997). Multitask Learning. *Machine Learning*, 28(1):41–75. p. 115
- Chatfield, C. (1989). *The Analysis of Time Series: An Introduction*. Chapman and Hall, London, 4th edition. pp. 82, 209, 214
- Choi, T. and Schervish, M. J. (2004). Posterior Consistency in Nonparametric Regression Problems Under Gaussian Process Priors. Technical Report 809, Department of Statistics, CMU. <http://www.stat.cmu.edu/tr/tr809/tr809.html>. p. 156
- Choudhuri, N., Ghosal, S., and Roy, A. (2005). Nonparametric Binary Regression Using a Gaussian Process Prior. Unpublished. <http://www4.stat.ncsu.edu/~sghosal/papers.html>. p. 156
- Chu, W. and Ghahramani, Z. (2005). Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research*, 6:1019–1041. p. 191
- Collins, M. and Duffy, N. (2002). Convolution Kernels for Natural Language. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*. MIT Press. p. 101
- Collobert, R. and Bengio, S. (2001). SVM Torch: Support Vector Machines for Large-Scale Regression Problems. *Journal of Machine Learning Research*, 1:143–160. <http://www.idiap.ch/~bengio/projects/SVM Torch.html>. pp. 69, 72
- Cornford, D., Nabney, I. T., and Williams, C. K. I. (2002). Modelling Frontal Discontinuities in Wind Fields. *Journal of Nonparameteric Statistics*, 14(1-2):43–58. p. 85
- Cox, D. D. (1984). Multivariate Smoothing Spline Functions. *SIAM Journal on Numerical Analysis*, 21(4):789–813. p. 156
- Cox, D. D. and O’Sullivan, F. (1990). Asymptotic Analysis of Penalized Likelihood and Related Estimators. *Annals of Statistics*, 18(4):1676–1695. p. 156

- Craven, P. and Wahba, G. (1979). Smoothing Noisy Data with Spline Functions. *Numer. Math.*, 31:377–403. p. 112
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York. pp. 30, 137, 190
- Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines*. Cambridge University Press. p. 141
- Cristianini, N., Shawe-Taylor, J., Elisseeff, A., and Kandola, J. (2002). On Kernel-Target Alignment. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*. MIT Press. p. 128
- Csató, L. (2002). *Gaussian Processes—Iterative Sparse Approximations*. PhD thesis, Aston University, UK. p. 179
- Csató, L. and Opper, M. (2002). Sparse On-Line Gaussian Processes. *Neural Computation*, 14(3):641–668. pp. 180, 185
- Csató, L., Opper, M., and Winther, O. (2002). TAP Gibbs Free Energy, Belief Propagation and Sparsity. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 657–663. MIT Press. pp. 180, 185
- Daley, R. (1991). *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, UK. p. 30
- David, H. A. (1970). *Order Statistics*. Wiley, New York. pp. 169, 170
- Davis, P. J. (1979). *Circulant Matrices*. Wiley, New York. p. 216
- Dawid, A. P. (1976). Properties of Diagnostic Data Distributions. *Biometrics*, 32:647–658. p. 34
- Dellaportas, P. and Stephens, D. A. (1995). Bayesian Analysis of Errors-in-Variables Regression Models. *Biometrics*, 51:1085–1095. p. 192
- Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer, New York. pp. 156, 166
- Diaconis, P. and Freedman, D. (1986). On the Consistency of Bayes Estimates. *Annals of Statistics*, 14(1):1–26. p. 156
- Diggle, P. J. (1990). *Time Series: A Biostatistical Introduction*. Clarendon Press, Oxford. pp. 214, 215
- Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based Geostatistics (with discussion). *Applied Statistics*, 47:299–350. p. 191
- Doob, J. L. (1944). The Elementary Gaussian Processes. *Annals of Mathematical Statistics*, 15(3):229–282. p. 212
- Doob, J. L. (1994). *Measure Theory*. Springer-Verlag, New York. p. 204
- Drineas, P. and Mahoney, M. W. (2005). On the Nyström Method for Approximating a Gram Matrix for Improved Kernel-Based Learning. Technical Report YALEU/DCS/TR-1319, Yale University. <http://cs-www.cs.yale.edu/homes/mmahoney>. p. 174

- Duchon, J. (1977). Splines Minimizing Rotation-Invariant Semi-norms in Sobolev Spaces. In Schempp, W. and Zeller, K., editors, *Constructive Theory of Functions of Several Variables*, pages 85–100. Springer-Verlag. p. 137
- Duda, R. O. and Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. John Wiley, New York. p. 146
- Edgeworth, F. Y. (1887). On Observations Relating to Several Quantities. *Hermathena*, 6:279–285. p. 146
- Faul, A. C. and Tipping, M. E. (2002). Analysis of Sparse Bayesian Learning. In Dietterich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 383–389, Cambridge, Massachusetts. MIT Press. p. 149
- Feldman, J. (1958). Equivalence and Perpendicularity of Gaussian Processes. *Pacific J. Math.*, 8:699–708. Erratum in *Pacific J. Math.* 9, 1295-1296 (1959). p. 157
- Ferrari Trecate, G., Williams, C. K. I., and Opper, M. (1999). Finite-dimensional Approximation of Gaussian Processes. In Kearns, M. S., Solla, S. A., and Cohn, D. A., editors, *Advances in Neural Information Processing Systems 11*, pages 218–224. MIT Press. p. 152
- Fine, S. and Scheinberg, K. (2002). Efficient SVM Training Using Low-Rank Kernel Representations. *Journal of Machine Learning Research*, 2(2):243–264. pp. 47, 174
- Fowlkes, C., Belongie, S., and Malik, J. (2001). Efficient Spatiotemporal Grouping Using the Nyström Method. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2001*. p. 172
- Freedman, D. (1999). On the Bernstein-Von Mises Theorem with Infinite-Dimensional Parameters. *Annals of Statistics*, 27(4):1119–1140. p. 156
- Frieze, A., Kannan, R., and Vempala, S. (1998). Fast Monte-Carlo Algorithms for Finding Low-Rank Approximations. In *39th Conference on the Foundations of Computer Science*, pages 370–378. p. 174
- Geisser, S. and Eddy, W. F. (1979). A Predictive Approach to Model Selection. *Journal of the American Statistical Association*, 74(365):153–160. p. 117
- Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6(6):721–741. p. 216
- Gibbs, M. N. (1997). *Bayesian Gaussian Processes for Regression and Classification*. PhD thesis, Department of Physics, University of Cambridge. p. 93
- Gibbs, M. N. and MacKay, D. J. C. (1997). Efficient Implementation of Gaussian Processes. Unpublished manuscript. Cavendish Laboratory, Cambridge, UK. <http://www.inference.phy.cam.ac.uk/mackay/BayesGP.html>. p. 181
- Gibbs, M. N. and MacKay, D. J. C. (2000). Variational Gaussian Process Classifiers. *IEEE Transactions on Neural Networks*, 11(6):1458–1464. p. 41
- Gihman, I. I. and Skorohod, A. V. (1974). *The Theory of Stochastic Processes*, volume 1. Springer Verlag, Berlin. p. 82

- Girard, A., Rasmussen, C. E., Quiñonero Candela, J., and Murray-Smith, R. (2003). Gaussian Process Priors With Uncertain Inputs: Application to Multiple-Step Ahead Time Series Forecasting. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*. MIT Press. p. 192
- Girosi, F. (1991). Models of Noise and Robust Estimates. Technical Report AI Memo 1287, MIT AI Laboratory. p. 146
- Girosi, F., Jones, M., and Poggio, T. (1995). Regularization Theory and Neural Networks Architectures. *Neural Computation*, 7(2):219–269. p. 25
- Goldberg, P. W., Williams, C. K. I., and Bishop, C. M. (1998). Regression with Input-dependent Noise: A Gaussian Process Treatment. In Jordan, M. I., Kearns, M. J., and Solla, S. A., editors, *Advances in Neural Information Processing Systems 10*. MIT Press, Cambridge, MA. p. 191
- Golub, G. H. and Van Loan, C. F. (1989). *Matrix Computations*. Johns Hopkins University Press, Baltimore. Second edition. pp. 172, 173, 174, 181, 202
- Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals, Series and Products*. Academic Press. Corrected and enlarged edition prepared by A. Jeffrey. pp. 98, 103
- Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, London. p. 138
- Grenander, U., Chow, Y., and Keenan, D. M. (1991). *Hands: A Pattern Theoretic Study of Biological Shapes*. Springer-Verlag, New York. pp. 214, 216
- Grimmett, G. R. and Stirzaker, D. R. (1992). *Probability and Random Processes*. Oxford University Press, Oxford, England, second edition. pp. 94, 213
- Grünwald, P. D. and Langford, J. (2004). Suboptimal Behaviour of Bayes and MDL in Classification Under Misspecification. In *Proc. Seventeenth Annual Conference on Computational Learning Theory (COLT 2004)*. p. 156
- Györfi, L., Kohler, M., Krzyżak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer, New York. p. 156
- Hajek, J. (1958). On a Property of Normal Distributions of Any Stochastic Process (In Russian). *Czechoslovak Math. J.*, 8:610–618. Translated in *Selected Trans. Math. Statist. Probab.* 1 245–252 (1961). Also available in *Collected Works of Jaroslav Hajek*, eds. M. Hušková, R. Beran, V. Dupač, Wiley, (1998). p. 157
- Hand, D. J., Mannila, H., and Smyth, P. (2001). *Principles of Data Mining*. MIT Press. p. 100
- Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York. p. 215
- Hansen, L. K., Liisberg, C., and Salamon, P. (1997). The Error-Reject Tradeoff. *Open Sys. & Information Dyn.*, 4:159–184. p. 36
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall. pp. 24, 25, 95
- Haussler, D. (1999). Convolution Kernels on Discrete Structures. Technical Report UCSC-CRL-99-10, Dept of Computer Science, University of California at Santa Cruz. p. 101

- Hawkins, D. L. (1989). Some Practical Problems in Implementing a Certain Sieve Estimator of the Gaussian Mean Function. *Communications in Statistics—Simulation and Computation*, 18(2):481–500. p. 97
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67. p. 11
- Hornik, K. (1993). Some New Results on Neural Network Approximation. *Neural Networks*, 6(8):1069–1072. p. 90
- Ihara, S. (1993). *Information Theory for Continuous Systems*. World Scientific, Singapore. p. 218
- Jaakkola, T. S., Diekhans, M., and Haussler, D. (2000). A Discriminative Framework for Detecting Remote Protein Homologies. *Journal of Computational Biology*, 7:95–114. pp. 101, 102, 104
- Jaakkola, T. S. and Haussler, D. (1999). Probabilistic Kernel Regression Models. In Heckerman, D. and Whittaker, J., editors, *Workshop on Artificial Intelligence and Statistics 7*. Morgan Kaufmann. p. 41
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., and Hinton, G. E. (1991). Adaptive Mixtures of Local Experts. *Neural Computation*, 3:79–87. p. 192
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions volume 2*. John Wiley and Sons, New York, second edition. p. 45
- Jones, D. R. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *J. Global Optimization*, 21:345–383. p. 193
- Jordan, M. I. (2005). An Introduction to Probabilistic Graphical Models. Draft book. p. 218
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press. p. 30
- Kailath, T. (1971). RKHS Approach to Detection and Estimation Problems—Part I: Deterministic Signals in Gaussian Noise. *IEEE Trans. Information Theory*, 17(5):530–549. p. 131
- Kammler, D. W. (2000). *A First Course in Fourier Analysis*. Prentice-Hall, Upper Saddle River, NJ. p. 208
- Kashyap, R. L. and Chellappa, R. (1981). Stochastic Models for Closed Boundary Analysis: Representation and Reconstruction. *IEEE Trans. on Information Theory*, 27(5):627–637. p. 216
- Keeling, C. D. and Whorf, T. P. (2004). Atmospheric CO<sub>2</sub> Records from Sites in the SIO Air Sampling Network. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn., U.S.A. p. 119
- Kent, J. T. and Mardia, K. V. (1994). The Link Between Kriging and Thin-plate Splines. In Kelly, F. P., editor, *Probability, Statistics and Optimization*, pages 325–339. Wiley. p. 137
- Kimeldorf, G. and Wahba, G. (1970). A Correspondence between Bayesian Estimation of Stochastic Processes and Smoothing by Splines. *Annals of Mathematical Statistics*, 41:495–502. p. 136
- Kimeldorf, G. and Wahba, G. (1971). Some Results on Tchebycheffian Spline Functions. *J. Mathematical Analysis and Applications*, 33(1):82–95. p. 132

- Kohn, R. and Ansley, C. F. (1987). A New Algorithm for Spline Smoothing based on Smoothing a Stochastic Process. *SIAM J. Sci. Stat. Comput.*, 8(1):33–48. p. 141
- Kolmogorov, A. N. (1941). Interpolation und Extrapolation von stationären zufälligen Folgen. *Izv. Akad. Nauk SSSR*, 5:3–14. p. 29
- König, H. (1986). *Eigenvalue Distribution of Compact Operators*. Birkhäuser. p. 96
- Kullback, S. (1959). *Information Theory and Statistics*. Dover, New York. pp. 158, 203
- Kuss, M. and Rasmussen, C. E. (2005). Assessing Approximations for Gaussian Process Classification. In Weiss, Y., Schölkopf, B., and Platt, J., editors, *Advances in Neural Information Processing Systems 18*. MIT Press. pp. 72, 73
- Lanckriet, G. R. G., Cristianini, N., Bartlett, P. L., El Ghaoui, L., and Jordan, M. I. (2004). Learning the Kernel Matrix with Semidefinite Programming. *Journal of Machine Learning Research*, 5(1):27–72. p. 128
- Lauritzen, S. L. (1981). Time Series Analysis in 1880: A Discussion of Contributions Made by T. N. Thiele. *International Statistical Review*, 49:319–333. p. 29
- Lawrence, N. (2004). Gaussian Process Latent Variable Models for Visualization of High Dimensional Data. In Thrun, S., Saul, L., and Schölkopf, B., editors, *Advances in Neural Information Processing Systems 16*, pages 329–336. MIT Press. p. 196
- Lawrence, N., Seeger, M., and Herbrich, R. (2003). Fast Sparse Gaussian Process Methods: The Informative Vector Machine. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 625–632. MIT Press. pp. 178, 185
- Leslie, C., Eskin, E., Weston, J., and Stafford Noble, W. (2003). Mismatch String Kernels for SVM Protein Classification. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*. MIT Press. pp. 100, 101, 104
- Lin, X., Wahba, G., Xiang, D., Gao, F., Klein, R., and Klein, B. (2000). Smoothing Spline ANOVA Models for Large Data Sets With Bernoulli Observations and the Randomized GACV. *Annals of Statistics*, 28:1570–1600. p. 185
- Lindley, D. V. (1985). *Making Decisions*. John Wiley and Sons, London, UK, second edition. p. 111
- Lodhi, H., Shawe-Taylor, J., Cristianini, N., and Watkins, C. J. C. H. (2001). Text Classification using String Kernels. In Leen, T. K., Diettrich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*. MIT Press. p. 101
- Luo, Z. and Wahba, G. (1997). Hybrid Adaptive Splines. *J. Amer. Statist. Assoc.*, 92:107–116. p. 176
- MacKay, D. J. C. (1992a). A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation*, 4(3):448–472. pp. 109, 166, 167
- MacKay, D. J. C. (1992b). Bayesian Interpolation. *Neural Computation*, 4(3):415–447. pp. xiii, xvi, 109
- MacKay, D. J. C. (1992c). Information-Based Objective Functions for Active Data Selection. *Neural Computation*, 4(4):590–604. p. 178

- MacKay, D. J. C. (1992d). The Evidence Framework Applied to Classification Networks. *Neural Computation*, 4(5):720–736. p. 45
- MacKay, D. J. C. (1998). Introduction to Gaussian Processes. In Bishop, C. M., editor, *Neural Networks and Machine Learning*. Springer-Verlag. pp. 84, 92
- MacKay, D. J. C. (1999). Comparison of Approximate Methods for Handling Hyperparameters. *Neural Computation*, 11(5):1035–1068. p. 110
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, Cambridge, UK. pp. xiv, 167
- Malzahn, D. and Opper, M. (2002). A Variational Approach to Learning Curves. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*. MIT Press. p. 161
- Mandelbrot, B. B. (1982). *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco. p. 137
- Mardia, K. V. and Marshall, R. J. (1984). Maximum Likelihood Estimation for Models of Residual Covariance in Spatial Regression. *Biometrika*, 71(1):135–146. p. 115
- Matérn, B. (1960). *Spatial Variation*. Meddelanden från Statens Skogsforskningsinstitut, 49, No.5. Almännas Förlaget, Stockholm. Second edition (1986), Springer-Verlag, Berlin. pp. 85, 87, 89
- Matheron, G. (1973). The Intrinsic Random Functions and Their Applications. *Advances in Applied Probability*, 5:439–468. p. 30
- Maxwell, J. C. (1850). Letter to Lewis Campbell; reproduced in L. Campbell and W. Garrett, *The Life of James Clerk Maxwell*, Macmillan, 1881. p. v
- McAllester, D. (2003). PAC-Bayesian Stochastic Model Selection. *Machine Learning*, 51(1):5–21. p. 164
- McCullagh, P. and Nelder, J. (1983). *Generalized Linear Models*. Chapman and Hall. pp. 37, 38, 138
- Meinguet, J. (1979). Multivariate Interpolation at Arbitrary Points Made Simple. *Journal of Applied Mathematics and Physics (ZAMP)*, 30:292–304. p. 137
- Meir, R. and Zhang, T. (2003). Generalization Error Bounds for Bayesian Mixture Algorithms. *Journal of Machine Learning Research*, 4(5):839–860. p. 164
- Micchelli, C. A. and Pontil, M. (2005). Kernels for Multi-task Learning. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 921–928. MIT Press. p. 190
- Micchelli, C. A. and Wahba, G. (1981). Design Problems for Optimal Surface Interpolation. In Ziegler, Z., editor, *Approximation Theory and Applications*, pages 329–348. Academic Press. p. 161
- Minka, T. P. (2001). *A Family of Algorithms for Approximate Bayesian Inference*. PhD thesis, Massachusetts Institute of Technology. pp. 41, 52
- Minka, T. P. (2003). A Comparison of Numerical Optimizers for Logistic Regression. <http://research.microsoft.com/~minka/papers/logreg>. p. 38



- Minka, T. P. and Picard, R. W. (1999). Learning How to Learn is Learning With Point Sets. <http://research.microsoft.com/~minka/papers/point-sets.html>. p. 116
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill, New York. pp. 2, 165
- Murray-Smith, R. and Girard, A. (2001). Gaussian Process priors with ARMA noise models. In *Irish Signals and Systems Conference*, pages 147–153, Maynooth. <http://www.dcs.gla.ac.uk/~rod/publications/MurGir01.pdf>. p. 191
- Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer, New York. Lecture Notes in Statistics 118. pp. xiii, 30, 90, 91, 106, 166, 167
- Neal, R. M. (1997). Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification. Technical Report 9702, Department of Statistics, University of Toronto. <http://www.cs.toronto.edu/~radford>. p. 191
- Neal, R. M. (1999). Regression and Classification using Gaussian Process Priors. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 6*, pages 475–501. Oxford University Press. (with discussion). pp. 41, 47
- Neal, R. M. (2001). Annealed Importance Sampling. *Statistics and Computing*, 11:125–139. p. 72
- Neumaier, A. (2005). Introduction to Global Optimization. <http://www.mat.univie.ac.at/~neum/glopt/intro.html>. p. 193
- O’Hagan, A. (1978). Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society B*, 40:1–42. (with discussion). pp. 28, 30, 94
- O’Hagan, A. (1991). Bayes-Hermite Quadrature. *Journal of Statistical Planning and Inference*, 29:245–260. pp. 193, 194
- O’Hagan, A., Kennedy, M. C., and Oakley, J. E. (1999). Uncertainty Analysis and other Inference Tools for Complex Computer Codes. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 6*, pages 503–524. Oxford University Press. (with discussion). p. 194
- Øksendal, B. (1985). *Stochastic Differential Equations*. Springer-Verlag, Berlin. p. 208
- Opper, M. and Vivarelli, F. (1999). General Bounds on Bayes Errors for Regression with Gaussian Processes. In Kearns, M. S., Solla, S. A., and Cohn, D. A., editors, *Advances in Neural Information Processing Systems 11*, pages 302–308. MIT Press. p. 160
- Opper, M. and Winther, O. (2000). Gaussian Processes for Classification: Mean-Field Algorithms. *Neural Computation*, 12(11):2655–2684. pp. 41, 44, 52, 127, 128
- O’Sullivan, F., Yandell, B. S., and Raynor, W. J. (1986). Automatic Smoothing of Regression Functions in Generalized Linear Models. *Journal of the American Statistical Association*, 81:96–103. pp. 132, 138
- Paciorek, C. and Schervish, M. J. (2004). Nonstationary Covariance Functions for Gaussian Process Regression. In Thrun, S., Saul, L., and Schölkopf, B., editors, *Advances in Neural Information Processing Systems 16*. MIT Press. pp. 93, 94
- Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York. Third Edition. pp. 79, 153, 154, 191, 211, 212

- Plaskota, L. (1996). *Noisy Information and Computational Complexity*. Cambridge University Press, Cambridge. pp. 161, 169
- Plate, T. A. (1999). Accuracy versus Interpretability in Flexible Modeling: Implementing a Tradeoff using Gaussian Process Models. *Behaviourmetrika*, 26(1):29–50. p. 95
- Platt, J. C. (1999). Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In Schölkopf, B., Burges, C. J. C., and Smola, A. J., editors, *Advances in Kernel Methods*, pages 185–208. MIT Press. p. 144
- Platt, J. C. (2000). Probabilities for SV Machines. In Smola, A., Bartlett, P., Schölkopf, B., and Schuurmans, D., editors, *Advances in Large Margin Classifiers*, pages 61–74. MIT Press. pp. 69, 70, 145, 147, 148
- Poggio, T. and Girosi, F. (1990). Networks for Approximation and Learning. *Proceedings of IEEE*, 78:1481–1497. pp. 89, 133, 134, 135, 147, 176
- Poggio, T., Voorhees, H., and Yuille, A. (1985). A Regularized Solution to Edge Detection. Technical Report AI Memo 833, MIT AI Laboratory. p. 154
- Pontil, M., Mukherjee, S., and Girosi, F. (1998). On the Noise Model of Support Vector Machine Regression. Technical Report AI Memo 1651, MIT AI Laboratory. p. 146
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, Second edition. pp. 96, 99, 201
- Quiñonero-Candela, J. (2004). *Learning with Uncertainty—Gaussian Processes and Relevance Vector Machines*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark. p. 177
- Rasmussen, C. E. (1996). *Evaluation of Gaussian Processes and Other Methods for Non-linear Regression*. PhD thesis, Dept. of Computer Science, University of Toronto. <http://www.kyb.mpg.de/publications/pss/ps2304.ps>. p. 30
- Rasmussen, C. E. (2003). Gaussian Processes to Speed up Hybrid Monte Carlo for Expensive Bayesian Integrals. In Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 7*, pages 651–659. Oxford University Press. p. 193
- Rasmussen, C. E. and Ghahramani, Z. (2001). Occam’s Razor. In Leen, T., Dietterich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages 294–300. MIT Press. p. 110
- Rasmussen, C. E. and Ghahramani, Z. (2002). Infinite Mixtures of Gaussian Process Experts. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*. MIT Press. p. 192
- Rasmussen, C. E. and Ghahramani, Z. (2003). Bayesian Monte Carlo. In Suzanna Becker, S. T. and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 489–496. MIT Press. p. 193
- Rasmussen, C. E. and Quiñonero-Candela, J. (2005). Healing the Relevance Vector Machine through Augmentation. In *Proc. 22nd International Conference on Machine Learning*. pp. 150, 176

- Rifkin, R. and Klautau, A. (2004). In Defense of One-Vs-All Classification. *Journal of Machine Learning Research*, 5:101–141. pp. 146, 147
- Ripley, B. (1981). *Spatial Statistics*. Wiley, New York. p. 30
- Ripley, B. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, UK. p. 35
- Ritter, K. (2000). *Average-Case Analysis of Numerical Problems*. Springer Verlag. pp. 159, 161, 169, 193
- Ritter, K., Wasilkowski, G. W., and Woźniakowski, H. (1995). Multivariate Integration and Approximation of Random Fields Satisfying Sacks-Ylvisaker Conditions. *Annals of Applied Probability*, 5:518–540. p. 97
- Rousseeuw, P. J. (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79:871–880. p. 146
- Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–435. pp. 16, 30
- Saitoh, S. (1988). *Theory of Reproducing Kernels and its Applications*. Longman, Harlow, England. p. 129
- Salton, G. and Buckley, C. (1988). Term-weighting Approaches in Automatic Text Retrieval. *Information Processing and Management*, 24:513–523. p. 100
- Sampson, P. D. and Guttorp, P. (1992). Nonparametric Estimation of Nonstationary Covariance Structure. *Journal of the American Statistical Association*, 87:108–119. p. 92
- Santner, T. J., Williams, B. J., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York. p. 30
- Saunders, C., Gammernan, A., and Vovk, V. (1998). Ridge Regression Learning Algorithm in Dual Variables. In Shavlik, J., editor, *Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998)*. Morgan Kaufmann. p. 30
- Saunders, C., Shawe-Taylor, J., and Vinokourov, A. (2003). String Kernels, Fisher Kernels and Finite State Automata. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*. MIT Press. p. 101
- Schoenberg, I. J. (1938). Metric Spaces and Positive Definite Functions. *Trans. American Mathematical Society*, 44(3):522–536. p. 86
- Schoenberg, I. J. (1964). Spline Functions and the Problem of Graduation. *Proc. Nat. Acad. Sci. USA*, 52:947–950. pp. 132, 138
- Schölkopf, B. and Smola, A. J. (2002). *Learning with Kernels*. MIT Press. pp. xvi, 73, 89, 90, 91, 129, 130, 133, 141, 144, 147, 173, 188, 195, 196
- Schölkopf, B., Smola, A. J., and Müller, K.-R. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, 10:1299–1319. p. 99

- Schwaighofer, A. and Tresp, V. (2003). Transductive and Inductive Methods for Approximate Gaussian Process Regression. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*. MIT Press. pp. 181, 184
- Scott, D. W. (1992). *Multivariate Density Estimation*. Wiley, New York. p. 25
- Seeger, M. (2000). Bayesian Model Selection for Support Vector Machines, Gaussian Processes and Other Kernel Classifiers. In Solla, S. A., Leen, T. K., and Müller, K.-R., editors, *Advances in Neural Information Processing Systems 12*. MIT Press, Cambridge, MA. pp. 41, 54, 145
- Seeger, M. (2002). PAC-Bayesian Generalisation Error Bounds for Gaussian Process Classification. *Journal of Machine Learning Research*, 3:322–269. pp. 161, 164, 165
- Seeger, M. (2003). *Bayesian Gaussian Process Models: PAC-Bayesian Generalisation Error Bounds and Sparse Approximations*. PhD thesis, School of Informatics, University of Edinburgh. <http://www.cs.berkeley.edu/~mseeger>. pp. 46, 145, 161, 162, 163, 179, 180, 186
- Seeger, M. (2005). Expectation Propagation for Exponential Families. <http://www.cs.berkeley.edu/~mseeger/papers/epexpfam.ps.gz>. p. 127
- Seeger, M. and Jordan, M. I. (2004). Sparse Gaussian Process Classification With Multiple Classes. Technical Report TR 661, Department of Statistics, University of California at Berkeley. p. 50
- Seeger, M., Williams, C. K. I., and Lawrence, N. (2003). Fast Forward Selection to Speed Up Sparse Gaussian Process Regression. In Bishop, C. and Frey, B. J., editors, *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*. Society for Artificial Intelligence and Statistics. pp. 178, 180
- Shawe-Taylor, J. and Williams, C. K. I. (2003). The Stability of Kernel Principal Components Analysis and its Relation to the Process Eigenspectrum. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*. MIT Press. p. 99
- Shepp, L. A. (1966). Radon-Nikodym Derivatives of Gaussian Measures. *Annals of Mathematical Statistics*, 37(2):321–354. p. 139
- Silverman, B. W. (1978). Density Ratios, Empirical Likelihood and Cot Death. *Applied Statistics*, 27(1):26–33. p. 138
- Silverman, B. W. (1984). Spline Smoothing: The Equivalent Variable Kernel Method. *Annals of Statistics*, 12(3):898–916. pp. 25, 153, 154
- Silverman, B. W. (1985). Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting (with discussion). *J. Roy. Stat. Soc. B*, 47(1):1–52. pp. 170, 175
- Simard, P., Victorri, B., Le Cun, Y., and Denker, J. (1992). Tangent Prop—A Formalism for Specifying Selected Invariances in an Adaptive Network. In Moody, J. E., Hanson, S. J., and Lippmann, R. P., editors, *Advances in Neural Information Processing Systems 4*, pages 895–903. Morgan Kaufmann. pp. 73, 195
- Smola, A. J. and Bartlett, P. L. (2001). Sparse Greedy Gaussian Process Regression. In Leen, T. K., Diettrich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages 619–625. MIT Press. p. 176

- Smola, A. J. and Schölkopf, B. (2000). Sparse Greedy Matrix Approximation for Machine Learning. In *Proceedings of the Seventeenth International Conference on Machine Learning*. Morgan Kaufmann. pp. 173, 174
- Solak, E., Murray-Smith, R., Leithead, W. E., Leith, D., and Rasmussen, C. E. (2003). Derivative Observations in Gaussian Process Models of Dynamic Systems. In Becker, S., S. T. and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 1033–1040. MIT Press. p. 191
- Sollich, P. (1999). Learning Curves for Gaussian Processes. In Kearns, M. S., Solla, S. A., and Cohn, D. A., editors, *Neural Information Processing Systems, Vol. 11*. MIT Press. pp. 160, 161
- Sollich, P. (2002). Bayesian Methods for Support Vector Machines: Evidence and Predictive Class Probabilities. *Machine Learning*, 46:21–52. pp. 145, 150, 161
- Sollich, P. and Williams, C. K. I. (2005). Using the Equivalent Kernel to Understand Gaussian Process Regression. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*. MIT Press. p. 154
- Stein, M. L. (1991). A Kernel Approximation to the Kriging Predictor of a Spatial Process. *Ann. Inst. Statist. Math*, 43(1):61–75. p. 154
- Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer-Verlag, New York. pp. 82, 83, 85, 86, 87, 115, 137, 157, 158, 161, 212
- Steinwart, I. (2005). Consistency of Support Vector Machines and Other Regularized Kernel Classifiers. *IEEE Trans. on Information Theory*, 51(1):128–142. p. 157
- Stitson, M. O., Gammernan, A., Vapnik, V. N., Vovk, V., Watkins, C. J. C. H., and Weston, J. (1999). Support Vector Regression with ANOVA Decomposition Kernels. In Schölkopf, B., Burges, C. J. C., and Smola, A. J., editors, *Advances in Kernel Methods*. MIT Press. p. 95
- Sundararajan, S. and Keerthi, S. S. (2001). Predictive Approaches for Choosing Hyperparameters in Gaussian Processes. *Neural Computation*, 13:1103–1118. p. 117
- Suykens, J. A. K. and Vanderwalle, J. (1999). Least Squares Support Vector Machines. *Neural Processing Letters*, 9:293–300. p. 147
- Szeliski, R. (1987). Regularization uses Fractal Priors. In *Proceedings of the 6th National Conference on Artificial Intelligence (AAAI-87)*. pp. 135, 137
- Teh, Y. W., Seeger, M., and Jordan, M. I. (2005). Semiparametric Latent Factor Models. In Cowell, R. G. and Ghahramani, Z., editors, *Proceedings of Tenth International Workshop on Artificial Intelligence and Statistics*, pages 333–340. Society for Artificial Intelligence and Statistics. p. 190
- Thomas-Agnan, C. (1996). Computing a Family of Reproducing Kernels for Statistical Applications. *Numerical Algorithms*, 13:21–32. p. 154
- Thompson, P. D. (1956). Optimum Smoothing of Two-Dimensional Fields. *Tellus*, 8:384–393. p. 30
- Tikhonov, A. N. (1963). Solution of Incorrectly Formulated Problems and the Regularization Method. *Soviet. Math. Dokl.*, 5:1035–1038. p. 133

- Tikhonov, A. N. and Arsenin, V. Y. (1977). *Solutions of Ill-posed Problems*. W. H. Winston, Washington, D.C. p. 133
- Tipping, M. E. (2001). Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1:211–244. p. 149
- Tipping, M. E. and Faul, A. C. (2003). Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. In Bishop, C. M. and Frey, B. J., editors, *Proceedings of Ninth International Workshop on Artificial Intelligence and Statistics*. Society for Artificial Intelligence and Statistics. p. 149
- Tresp, V. (2000). A Bayesian Committee Machine. *Neural Computation*, 12(11):2719–2741. pp. 180, 181, 185, 187
- Tsuda, K., Kawanabe, M., Rätsch, G., Sonnenburg, S., and Müller, K.-R. (2002). A New Discriminative Kernel from Probabilistic Models. *Neural Computation*, 14(10):2397–2414. p. 102
- Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the Theory of Brownian Motion. *Phys. Rev.*, 36:823–841. pp. 86, 212
- Valiant, L. G. (1984). A Theory of the Learnable. *Communications of the ACM*, 27(11):1134–1142. p. 161
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer Verlag, New York. pp. 36, 141, 181
- Vapnik, V. N. (1998). *Statistical Learning Theory*. John Wiley and Sons. p. 140
- Vijayakumar, S., D’Souza, A., and Schaal, S. (2005). Incremental Online Learning in High Dimensions. Accepted for publication in *Neural Computation*. pp. 22, 24
- Vijayakumar, S., D’Souza, A., Shibata, T., Conradt, J., and Schaal, S. (2002). Statistical Learning for Humanoid Robots. *Autonomous Robot*, 12(1):55–69. p. 22
- Vijayakumar, S. and Schaal, S. (2000). LWPR: An  $O(n)$  Algorithm for Incremental Real Time Learning in High Dimensional Space. In *Proc. of the Seventeenth International Conference on Machine Learning (ICML 2000)*, pages 1079–1086. p. 22
- Vishwanathan, S. V. N. and Smola, A. J. (2003). Fast Kernels for String and Tree Matching. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*. MIT Press. p. 101
- Vivarelli, F. and Williams, C. K. I. (1999). Discovering Hidden Features with Gaussian Processes Regression. In Kearns, M. S., Solla, S. A., and Cohn, D. A., editors, *Advances in Neural Information Processing Systems 11*. MIT Press. p. 89
- von Mises, R. (1964). *Mathematical Theory of Probability and Statistics*. Academic Press. p. 200
- Wahba, G. (1978). Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression. *Journal of the Royal Statistical Society B*, 40(3):364–372. p. 139
- Wahba, G. (1985). A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem. *Annals of Statistics*, 13:1378–1402. p. 29

- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, Philadelphia, PA. CBMS-NSF Regional Conference series in applied mathematics. pp. 95, 112, 117, 118, 129, 131, 137, 138, 157, 176
- Wahba, G., Johnson, D. R., Gao, F., and Gong, J. (1995). Adaptive Tuning of Numerical Weather Prediction Models: Randomized GCV in Three-and Four-Dimensional Data Assimilation. *Monthly Weather Review*, 123:3358–3369. p. 181
- Watkins, C. J. C. H. (1999). Dynamic Alignment Kernels. Technical Report CSD-TR-98-11, Dept of Computer Science, Royal Holloway, University of London. p. 101
- Watkins, C. J. C. H. (2000). Dynamic Alignment Kernels. In Smola, A. J., Bartlett, P. L., and Schölkopf, B., editors, *Advances in Large Margin Classifiers*, pages 39–50. MIT Press, Cambridge, MA. p. 100
- Wegman, E. J. (1982). Reproducing Kernel Hilbert Spaces. In Kotz, S. and Johnson, N. L., editors, *Encyclopedia of Statistical Sciences*, volume 8, pages 81–84. Wiley, New York. pp. 129, 130
- Weinert, H. L., editor (1982). *Reproducing Kernel Hilbert Spaces*. Hutchinson Ross, Stroudsburg, Pennsylvania. p. 129
- Wendland, H. (2005). *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press. p. 88
- Whittle, P. (1963). *Prediction and Regulation by Linear Least-square Methods*. English Universities Press. pp. 30, 216
- Widom, H. (1963). Asymptotic Behavior of the Eigenvalues of Certain Integral Equations. *Trans. of the American Mathematical Society*, 109(2):278–295. p. 97
- Widom, H. (1964). Asymptotic Behavior of the Eigenvalues of Certain Integral Equations II. *Archive for Rational Mechanics and Analysis*, 17:215–229. p. 97
- Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series*. MIT Press, Cambridge, Mass. p. 29
- Williams, C. K. I. (1998). Computation with Infinite Neural Networks. *Neural Computation*, 10(5):1203–1216. p. 91
- Williams, C. K. I. and Barber, D. (1998). Bayesian Classification with Gaussian Processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1342–1351. pp. 41, 45, 48, 49
- Williams, C. K. I. and Rasmussen, C. E. (1996). Gaussian Processes for Regression. In Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems 8*, pages 514–520. MIT Press. pp. 30, 107
- Williams, C. K. I., Rasmussen, C. E., Schwaighofer, A., and Tresp, V. (2002). Observations on the Nyström Method for Gaussian Process Prediction. Technical report, University of Edinburgh. <http://www.dai.ed.ac.uk/homes/ckiw/online-pubs.html>. p. 177
- Williams, C. K. I. and Seeger, M. (2001). Using the Nyström Method to Speed Up Kernel Machines. In Leen, T. K., Diettrich, T. G., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages 682–688. MIT Press. pp. 173, 177

- Williams, C. K. I. and Vivarelli, F. (2000). Upper and Lower Bounds on the Learning Curve for Gaussian Processes. *Machine Learning*, 40:77–102. pp. 31, 161, 168
- Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer, Berlin. p. 219
- Wong, E. (1971). *Stochastic Processes in Information and Dynamical Systems*. McGraw-Hill, New York. p. 218
- Wood, S. and Kohn, R. (1998). A Bayesian Approach to Robust Binary Nonparametric Regression. *J. American Statistical Association*, 93(441):203–213. p. 45
- Yaglom, A. M. (1987). *Correlation Theory of Stationary and Related Random Functions Volume I: Basic Results*. Springer Verlag. p. 89
- Yang, C., Duraiswami, R., and David, L. (2005). Efficient Kernel Machines Using the Improved Fast Gauss Transform. In Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*. MIT Press. p. 182
- Ylvisaker, D. (1975). Designs on Random Fields. In Srivastava, J. N., editor, *A Survey of Statistical Design and Linear Models*, pages 593–608. North-Holland. p. 159
- Yuille, A. and Grzywacz, N. M. (1989). A Mathematical Analysis of Motion Coherence Theory. *International Journal of Computer Vision*, 3:155–175. p. 134
- Zhang, T. (2004). Statistical Behaviour and Consistency of Classification Methods based on Convex Risk Minimization (with discussion). *Annals of Statistics*, 32(1):56–85. p. 157
- Zhu, H., Williams, C. K. I., Rohwer, R. J., and Morciniec, M. (1998). Gaussian Regression and Optimal Finite Dimensional Linear Models. In Bishop, C. M., editor, *Neural Networks and Machine Learning*. Springer-Verlag, Berlin. p. 97
- Zhu, J. and Hastie, T. J. (2002). Kernel Logistic Regression and the Import Vector Machine. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Information Processing Systems 14*, pages 1081–1088. MIT Press. p. 185