Kernels for Vector-Valued Functions: A Review

Kernels for Vector-Valued Functions: A Review

Mauricio A. Álvarez

Universidad Tecnológica de Pereira
Pereira 660003, Colombia
malvarez@utp.edu.co

Lorenzo Rosasco

Italy and Massachusetts Institute of Technology
Cambridge 02138, USA
lrosasco@mit.edu

Neil D. Lawrence

University of Sheffield
Sheffield, UK
N.Lawrence@dcs.sheffield.ac.uk

the essence of knowledge

Boston – Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is M. A. Álvarez, L. Rosasco and N. D. Lawrence, Kernels for Vector-Valued Functions: A Review, Foundation and Trends® in Machine Learning, vol 4, no 3, pp 195–266, 2011.

ISBN: 978-1-60198-558-3

Foundations and Trends® in Machine Learning

Volume 4 Issue 3, 2011

Editorial Scope

Foundations and Trends® in Machine Learning will publish survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2011, Volume 4, 4 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends® in
Machine Learning
Vol. 4, No. 3 (2011) 195–266

DOI: 10.1561/2200000036

Kernels for Vector-Valued Functions: A Review

Mauricio A. Álvarez[1], Lorenzo Rosasco[2] and Neil D. Lawrence[3]

[1] *Department of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira 660003, Colombia, malvarez@utp.edu.co*
[2] *Istituto Italiano di Tecnologia, Italy and Massachusetts Institute of Technology, 43 Vassar St, Cambridge 02138, USA, lrosasco@mit.edu*
[3] *Department of Computer Science, University of Sheffield and The Sheffield Institute for Translational Neuroscience, Sheffield, S1 4DP, UK, N.Lawrence@dcs.sheffield.ac.uk*

Abstract

Kernel methods are among the most popular techniques in machine learning. From a regularization perspective they play a central role in regularization theory as they provide a natural choice for the hypotheses space and the regularization functional through the notion of reproducing kernel Hilbert spaces. From a probabilistic perspective they are the key in the context of Gaussian processes, where the kernel function is known as the covariance function. Traditionally, kernel methods have been used in supervised learning problems with scalar outputs and indeed there has been a considerable amount of work devoted to designing and learning kernels. More recently there

has been an increasing interest in methods that deal with multiple outputs, motivated partially by frameworks like multitask learning. In this monograph, we review different methods to design or learn valid kernel functions for multiple outputs, paying particular attention to the connection between probabilistic and functional methods.

Contents

1

Introduction

Many modern applications of machine learning require solving several decision making or prediction problems and exploiting dependencies between the problems is often the key to obtain better results and coping with a lack of data (to solve a problem we can *borrow strength* from a distinct but related problem).

In *sensor networks*, for example, missing signals from certain sensors may be predicted by exploiting their correlation with observed signals acquired from other sensors [72]. In *geostatistics*, predicting the concentration of heavy pollutant metals, which are expensive to measure, can be done using inexpensive and oversampled variables as a proxy [37]. In *computer graphics*, a common theme is the animation and simulation of physically plausible humanoid motion. Given a set of poses that delineate a particular movement (for example, walking), we are faced with the task of completing a sequence by filling in the missing frames with natural-looking poses. Human movement exhibits a high degree of correlation. Consider, for example, the way we walk. When moving the right leg forward, we unconsciously prepare the left leg, which is currently touching the ground, to start moving as soon as the right leg reaches the floor. At the same time, our hands move synchronously with our legs. We can exploit these implicit correlations for predicting new poses and for generating new natural-looking walking sequences [106]. In *text categorization*, one document can be assigned to multiple topics or have multiple labels [50]. In all the examples above, the simplest approach ignores the potential correlation among the different output components of the problem and employ models that make predictions individually for each output. However, these examples suggest a different approach through a joint prediction exploiting the interaction between the different components to improve on individual predictions. Within the machine learning community this type of modeling is often broadly referred to as *multitask learning*. Again the key idea is that information shared between different tasks can lead to improved performance in comparison to learning the same tasks individually. These ideas are related to *transfer learning* [12, 20, 74, 97], a term which refers to systems that learn by transferring knowledge between different domains, for example: "what can we learn about running through seeing walking?"

More formally, the classical supervised learning problem requires estimating the output for any given input $\mathbf{x}_*$; an estimator $f_*(\mathbf{x}_*)$ is built on the basis of a training set consisting of N input–output pairs $S = (\mathbf{X}, \mathbf{Y}) = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$. The input space $\mathcal{X}$ is usually a space of vectors, while the output space is a space of *scalars*. In multiple output learning (MOL) the output space is a space of *vectors*; the estimator is now a *vector-valued function* $\mathbf{f}$. Indeed, this situation can also be described as the problem of solving D distinct classical supervised problems, where each problem is described by one of the components $f_1, \ldots, f_D$ of $\mathbf{f}$. As mentioned before, the key idea is to work under the assumption that the problems are in some way related. The idea is then to exploit the relation among the problems to improve upon solving each problem separately.

The goal of this survey is twofold. First, we aim at discussing recent results in multi-output/multitask learning based on kernel methods and Gaussian processes providing an account of the state of the art in the field. Second, we analyze systematically the connections between Bayesian and regularization (frequentist) approaches. Indeed, related techniques have been proposed from different perspectives and drawing clearer connections can boost advances in the field, while fostering collaborations between different communities.

The plan of the monograph follows. In Section 2 we give a brief review of the main ideas underlying kernel methods for scalar learning, introducing the concepts of regularization in reproducing kernel Hilbert spaces and Gaussian processes. In Section 3 we describe how similar concepts extend to the context of vector-valued functions and discuss different settings that can be considered. In Sections 4 and 5 we discuss approaches to constructing multiple output kernels, drawing connections between the Bayesian and regularization frameworks. The parameter estimation problem and the computational complexity problem are both described in Section 6. In Section 7 we discuss some potential applications that can be seen as multi-output learning. Finally we conclude in Section 8 with some remarks and discussion.

References

[1] M. A. Álvarez, "Convolved Gaussian process priors for multivariate regression with applications to dynamical systems," PhD thesis, School of Computer Science, University of Manchester, Manchester, UK, 2011.

[2] M. A. Álvarez and N. D. Lawrence, "Sparse convolved Gaussian processes for multi-output regression," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 21, (D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds.), pp. 57–64, Cambridge, MA: MIT Press, 2009.

[3] M. A. Álvarez, D. Luengo, and N. D. Lawrence, "Latent Force Models," in *Proceedings of the International Conference on Artificial Intelligence and Statistics*, (D. van Dyk and M. Welling, eds.), pp. 9–16, Clearwater Beach, Florida, 16–18 April 2009. JMLR W&CP 5.

[4] A. Argyriou, T. Evgeniou, and M. Pontil, "Convex multi-task feature learning," *Machine Learning*, vol. 73, no. 3, pp. 243–272, 2008.

[5] A. Argyriou, A. Maurer, and M. Pontil, "An algorithm for transfer learning in a heterogeneous environment," in *Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databse–Part I (ECML/PKDD)*, pp. 71–85, 2008.

[6] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.

[7] L. Baldassarre, L. Rosasco, A. Barla, and A. Verri, "Multi-output learning via spectral filtering," Technical Report, Massachusetts Institute of Technology, MIT-CSAIL-TR-2011-004, CBCL-296, 2011.

[8] R. P. Barry and J. M. Ver Hoef, "Blackbox kriging: spatial prediction without specifying variogram models," *Journal of Agricultural, Biological and Environmental Statistics*, vol. 1, no. 3, pp. 297–322, 1996.

[9] M. J. Bayarri, J. O. Berger, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, C.-H. Lin, and J. Tu, "Predicting vehicle crashworthiness: Validation of computer models for functional and hierarchical data," *Journal of the American Statistical Association*, vol. 104, no. 487, pp. 929–943, 2009.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*. Information Science and Statistics, Springer, 2006.

[11] E. Bonilla, S. Guo, and S. Sanner, "Gaussian process preference elicitation," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 24, pp. 262–270, Cambridge, MA: MIT Press, 2011.

[12] E. V. Bonilla, K. M. Chai, and C. K. I. Williams, "Multi-task Gaussian process prediction," in *Advances in Neural Information Processing Systems (NIPS)*vol. 20, (J. C. Platt, D. Koller, Y. Singer, and S. Roweis, eds.), Cambridge, MA: MIT Press, 2008.

[13] P. Boyle, "Gaussian processes for regression and optimisation," PhD thesis, Victoria University of Wellington, Wellington, New Zealand, 2007.

[14] P. Boyle and M. Frean, "Dependent Gaussian processes," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 17, (L. Saul, Y. Weiss, and L. Bouttou, eds.), pp. 217–224, Cambridge, MA: MIT Press, 2005.

[15] P. Boyle and M. Frean, "Multiple output Gaussian process regression," Technical Report CS-TR-05/2, School of Mathematical and Computing Sciences, Victoria University, New Zealand, 2005.

[16] C. A. Calder, "A dynamic process convolution approach to modeling ambient particulate matter concentrations," *Environmetrics*, vol. 19, pp. 39–48, 2008.

[17] C. A. Calder and N. Cressie, "Some topics in convolution-based spatial modeling," in *Proceedings of the 56th Session of the International Statistics Institute (ISI)*, August 2007.

[18] A. Caponnetto, C. Micchelli, M. Pontil, and Y. Ying, "Universal kernels for multi-task learning," *Journal of Machine Learning Research*, vol. 9, pp. 1615–1646, 2008.

[19] C. Carmeli, E. De Vito, and A. Toigo, "Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem," *Analysis and Applications (Singapore)*, vol. 4, no. 4, pp. 377–408, 2006.

[20] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, pp. 41–75, 1997.

[21] K. M. Chai, "Generalization errors and learning curves for regression with multi-task Gaussian processes," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 22, (Y. Bengio, D. Schuurmans, J. Laferty, C. Williams, and A. Culotta, eds.), pp. 279–287, Cambridge, MA: MIT Press, 2010.

[22] K. M. A. Chai, C. K. I. Williams, S. Klanke, and S. Vijayakumar, "Multitask Gaussian process learning of robot inverse dynamics," in *Advances in Neural Information Processing Systems (NIPS)*, (D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds.), pp. 265–272, Cambridge, MA: MIT Press, 2009.

[23] S. Conti and A. O'Hagan, "Bayesian emulation of complex multi-output and dynamic computer models," *Journal of Statistical Planning and Inference*, vol. 140, no. 3, pp. 640–651, 2010.

[24] N. A. C. Cressie, *Statistics for Spatial Data*. USA: John Wiley & Sons (Revised edition), 1993.

[25] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society (N.S.)*, vol. 39, no. 1, pp. 1–49, (electronic) 2002.

[26] E. De Vito, L. Rosasco, A. Caponnetto, M. Piana, and A. Verri, "Some properties of regularized kernel methods," *Journal of Machine Learning Research*, vol. 5, pp. 1363–1390, 2004.

[27] T. Evgeniou, C. A. Micchelli, and M. Pontil, "Learning multiple tasks with kernel methods," *Journal of Machine Learning Research*, vol. 6, pp. 615–637, 2005.

[28] T. Evgeniou, C. A. Micchelli, and M. Pontil, "Learning multiple tasks with kernel methods," *Journal of Machine Learning Research*, vol. 6, pp. 615–637, 2005.

[29] T. E. Fricker, J. E. Oakley, N. D. Sims, and K. Worden, "Probabilistic uncertainty analysis of an FRF of a structure using a Gaussian process emulator," *Mechanical Systems and Signal Processing*, vol. 25, no. 8, pp. 2962–2975, 2011.

[30] M. Fuentes, "Interpolation of nonstationary air pollution processes: A spatial spectral approach," *Statistical Modelling*, vol. 2, pp. 281–298, 2002.

[31] M. Fuentes, "Spectral methods for nonstationary spatial processes," *Biometrika*, vol. 89, no. 1, pp. 197–210, 2002.

[32] E. Fuselier Jr, "Refined error estimates for matrix-valued radial basis functions," PhD thesis, Texas A&M University, 2006.

[33] P. Gao, A. Honkela, M. Rattray, and N. D. Lawrence, "Gaussian process modelling of latent chemical species: Applications to inferring transcription factor activities," *Bioinformatics*, vol. 24, pp. i70–i75, 2008.

[34] A. E. Gelfand, A. M. Schmidt, S. Banerjee, and C. Sirmans, "Nonstationary multivariate process modeling through spatially varying coregionalization," *TEST*, vol. 13, no. 2, pp. 263–312, 2004.

[35] F. Girosi and T. Poggio, "Networks and the best approximation property," *Biological Cybernetics*, vol. 63, pp. 169–176, 1989.

[36] T. Gneiting, Z. Sasvári, and M. Schlather, "Analogies and correspondences between variograms and covariance functions," *Advances in Applied Probability*, vol. 33, no. 3, pp. 617–630, 2001.

[37] P. Goovaerts, *Geostatistics For Natural Resources Evaluation*. USA: Oxford University Press, 1997.

[38] M. Goulard and M. Voltz, "Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix," *Mathematical Geology*, vol. 24, no. 3, pp. 269–286, 1992.

[39] J. A. V. Guzmán, A. Warrick, and D. E. Myers, "Coregionalization by linear combination of nonorthogonal components," *Mathematical Geology*, vol. 34, no. 4, pp. 405–419, 2002.

[40] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, Second ed., 2009.

[41] J. D. Helterbrand and N. Cressie, "Universal cokriging under intrinsic coregionalization," *Mathematical Geology*, vol. 26, no. 2, pp. 205–226, 1994.

[42] D. Higdon, J. Gattiker, B. Williams, and M. Rightley, "Computer model calibration using high dimensional output," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 570–583, 2008.

[43] D. M. Higdon, "A process-convolution approach to modeling temperatures in the North Atlantic Ocean," *Journal of Ecological and Environmental Statistics*, vol. 5, pp. 173–190, 1998.

[44] D. M. Higdon, "Space and space-time modelling using process convolutions," in *Quantitative methods for current environmental issues*, (C. Anderson, V. Barnett, P. Chatwin, and A. El-Shaarawi, eds.), pp. 37–56, Springer-Verlag, 2002.

[45] D. M. Higdon, J. Swall, and J. Kern, "Non-stationary spatial modeling," in *Bayesian Statistics 6*, (J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds.), pp. 761–768, Oxford University Press, 1998.

[46] L. Hörmander, *The Analysis of Linear Partial Differential Operators I*. Berlin Hiedelberg: Springer-Verlag, First ed., 1983.

[47] L. Jacob, F. Bach, and J. Vert, "Clustered multi-task learning: A convex formulation," in *Advances in Neural Information Processing Systems (NIPS)*, Curran Associates, Inc, 2008.

[48] L. Jacob, F. Bach, and J.-P. Vert, "Clustered multi-task learning: A convex formulation," in *Advances in Neural Information Processing Systems (NIPS) 21*, pp. 745–752, 2008.

[49] A. G. Journel and C. J. Huijbregts, *Mining Geostatistics*. London: Academic Press, 1978.

[50] H. Kazawa, T. Izumitani, H. Taira, and E. Maeda, "Maximal margin labeling for multi-topic text categorization," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 17, (L. Saul, Y. Weiss, and L. Bouttou, eds.), pp. 649–656, Cambridge, MA: MIT Press, 2005.

[51] G. Kimeldorf and G. Wahba, "A correspondence between Bayesian estimation of stochastic processes and smoothing by splines," *Annals of Mathematical Statistics*, vol. 41, pp. 495–502, 1970.

[52] D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds., *Advances in Neural Information Processing Systems (NIPS)*. vol. 21, Cambridge, MA: MIT Press, 2009.

[53] H. Künsch, A. Papritz, and F. Bassi, "Generalized cross-covariances and their estimation," *Mathematical Geology*, vol. 29, no. 6, pp. 779–799, 1997.

[54] R. M. Lark and A. Papritz, "Fitting a linear model of coregionalization for soil properties using simulated annealing," *Geoderma*, vol. 115, pp. 245–260, 2003.

[55] N. D. Lawrence and J. C. Platt, "Learning to learn with the informative vector machine," in *Proceedings of the International Conference on Machine Learning (ICML 2004)*, pp. 512–519, 2004.

[56] N. D. Lawrence, G. Sanguinetti, and M. Rattray, "Modelling transcriptional regulation using Gaussian Processes," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 19, (B. Schölkopf, J. C. Platt, and T. Hofmann, eds.), pp. 785–792, Cambridge, MA: MIT Press, 2007.

[57] N. D. Lawrence, M. Seeger, and R. Herbrich, "Fast sparse Gaussian process methods: The informative vector machine," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 15, (S. Becker, S. Thrun, and K. Obermayer, eds.), pp. 625–632, Cambridge, MA: MIT Press, 2003.

[58] F. Liang, K. Mao, M. Liao, S. Mukherjee, and M. West, "Non-parametric Bayesian kernel models," Department of Statistical Science, Duke University, Discussion Paper 07-10. (Submitted for publication), 2009.

[59] S. Lowitzsch, "A density theorem for matrix-valued radial basis functions," *Numerical Algorithms*, vol. 39, no. 1, pp. 253–256, 2005.

[60] I. Macêdo and R. Castro, "Learning divergence-free and curl-free vector fields with matrix-valued kernels," Technical Report, Instituto Nacional de Matematica Pura e Aplicada, 2008.

[61] A. Majumdar and A. E. Gelfand, "Multivariate spatial modeling for geostatistical data using convolved covariance functions," *Mathematical Geology*, vol. 39, no. 2, pp. 225–244, 2007.

[62] G. Matheron, "The intrinsic random functions and their applications," *Advances in Applied Probability*, vol. 5, no. 3, pp. 439–468, 1973.

[63] J. McFarland, S. Mahadevan, V. Romero, and L. Swiler, "Calibration and Uncertainty Analysis for Computer Simulations with Multivariate Output," *American Institute of Aeronautics and Astronautics Journal*, vol. 46, no. 5, pp. 1253–1265, 2008.

[64] C. Micchelli and M. Pontil, "On learning vector-valued functions," *Neural Computation*, vol. 17, pp. 177–204, 2005.

[65] C. A. Micchelli and M. Pontil, "Kernels for multi-task learning," in *Advances in Neural Information Processing Systems (Advances in Neural Information Processing Systems (NIPS))*, MIT Press, 2004.

[66] T. P. Minka and R. W. Picard, "Learning how to learn is learning with point sets," Revised version 1999 available at http://research.microsoft.com/en-us/um/people/minka/papers/point-sets.html, 1999.

[67] R. Murray-Smith and B. A. Pearlmutter, "Transformation of Gaussian process priors," in *Deterministic and Statistical Methods in Machine Learning*, (J. Winkler, M. Niranjan, and N. Lawrence, eds.), pp. 110–123, LNAI 3635, Springer-Verlag, 2005.

[68] F. Narcowich and J. Ward, "Generalized Hermite interpolation via matrix-valued conditionally positive definite functions," *Mathematics of Computation*, vol. 63, no. 208, pp. 661–687, 1994.

[69] G. Obozinski, B. Taskar, and M. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Statistics and Computing*, vol. 20, no. 2, pp. 231–252, 2010.

[70] A. O'Hagan, "Bayesian analysis of computer code outputs: A tutorial," *Reliability Engineering and System Safety*, vol. 91, pp. 1290–1300, 2006.

[71] M. A. Osborne and S. J. Roberts, "Gaussian processes for prediction," Technical Report, Department of Engineering Science, University of Oxford, 2007.

[72] M. A. Osborne, A. Rogers, S. D. Ramchurn, S. J. Roberts, and N. R. Jennings, "Towards real-time information processing of sensor network data using computationally efficient multi-output Gaussian processes," in *Proceedings of the International Conference on Information Processing in Sensor Networks (IPSN 2008)*, 2008.

[73] C. J. Paciorek and M. J. Schervish, "Nonstationary covariance functions for Gaussian process regression," in *Advances in Neural Information Processing Systems (NIPS) 16*, (S. Thrun, L. Saul, and B. Schölkopf, eds.), Cambridge, MA: MIT Press, 2004.

[74] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, October 2010.

[75] A. Papritz, H. Künsch, and R. Webster, "On the Pseudo cross-variogram," *Mathematical Geology*, vol. 25, no. 8, pp. 1015–1026, 1993.

[76] B. Pelletier, P. Dutilleul, G. Larocque, and J. W. Fyles, "Fitting the linear model of coregionalization by generalized least squares," *Mathematical Geology*, vol. 36, no. 3, pp. 323–343, 2004.

[77] N. S. Pillai, Q. Wu, F. Liang, S. Mukherjee, and R. L. Wolpert, "Characterizing the function space for Bayesian kernel models," *Journal of Machine Learning Research*, vol. 8, pp. 1769–1797, 2007.

[78] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1481–1497, 1990.

[79] P. Z. G. Qian, H. Wu, and C. F. J. Wu, "Gaussian process models for computer experiments with qualitative and quantitative factors," *Technometrics*, vol. 50, no. 3, pp. 383–396, 2008.

[80] J. Quiñonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *Journal of Machine Learning Research*, vol. 6, pp. 1939–1959, 2005.

[81] J. Quiñonero-Candela and C. E. Rasmussen, "Analysis of some methods for reduced rank Gaussian process regression," in *Lecture Notes in Computer Science*, (R. Murray-Smith and R. Shorten, eds.), pp. 98–127, Springer, 2005.

[82] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.

[83] J. Rougier, "Efficient emulators for multivariate deterministic functions," *Journal of Computational and Graphical Statistics*, vol. 17, no. 4, pp. 827–834, 2008.

[84] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. Springer, First ed., 2003.

[85] L. Saul, Y. Weiss, and L. Bouttou, eds., *Advances in Neural Information Processing Systems (NIPS)*. vol. 17, Cambridge, MA: MIT Press, 2005.

[86] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. USA: The MIT Press, 2002.

[87] B. Schšlkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proceedings of the Annual Conference on Computational Learning Theory (COLT)*, pp. 416–426, 2001.

[88] L. Schwartz, "Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants)," *Journal d'Analyse Mathématique*, vol. 13, pp. 115–256, 1964.

[89] M. Seeger and M. I. Jordan, "Sparse Gaussian Process Classification With Multiple Classes," Technical Report 661, Department of Statistics, University of California at Berkeley, 2004.

[90] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. New York, NY, USA: Cambridge University Press, 2004.

[91] D. Sheldon, "Graphical multi-task learning," Technical Report, Cornell University, Preprint, 2008.

[92] J. Q. Shi, R. Murray-Smith, D. Titterington, and B. Pearlmutter, "Learning with large data sets using filtered Gaussian process priors," in *Proceedings of the Hamilton Summer School on Switching and Learning in Feedback Systems*, (R. Murray-Smith and R. Shorten, eds.), pp. 128–139, LNCS 3355, Springer-Verlag, 2005.

[93] G. Skolidis and G. Sanguinetti, "Bayesian multitask classification With Gaussian process priors," *IEEE Transactions on Neural Networks*, vol. 22, no. 12, pp. 2011–2021, 2011.

[94] O. Stegle, C. Lippert, J. Mooij, N. Lawrence, and K. Borgwardt, "Learning sparse inverse covariance matrices in the presence of confounders," in *Neural Information Processing Systems*, 2011.

[95] M. L. Stein, *Interpolation of Spatial Data*. Springer-Verlag New York, Inc., First ed., 1999.

[96] Y. W. Teh, M. Seeger, and M. I. Jordan, "Semiparametric latent factor models," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS 10)*, (R. G. Cowell and Z. Ghahramani, eds.), pp. 333–340, Barbados: Society for Artificial Intelligence and Statistics, 6–8 January 2005.

[97] S. Thrun, "Is learning the n-thing any easier than learning the first?," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 08, (D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, eds.), pp. 640–646, Cambridge, MA: MIT Press, 1996.

[98] A. Tikhonov and V. Arsenin, *Solutions of Ill Posed Problems*. Washington, D.C.: W. H. Winston, 1977.

[99] M. Titsias, N. D. Lawrence, and M. Rattray, "Efficient sampling for Gaussian process inference using control variables," in *Advances in Neural Information Processing Systems (NIPS)*, (D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, eds.), pp. 1681–1688, Cambridge, MA: MIT Press, 2009.

[100] V. N. Vapnik, *Statistical Learning Theory*. (Adaptive and Learning Systems for Signal Processing, Communications, and Control), New York: John Wiley & Sons Inc., 1998. A Wiley-Interscience Publication.

[101] E. Vazquez and E. Walter, "Multi-output support vector regression," in *13th IFAC Symposium on System Identification (SYSID)*, 2003.

[102] J. M. Ver Hoef and R. P. Barry, "Constructing and fitting models for cokriging and multivariable spatial prediction," *Journal of Statistical Planning and Inference*, vol. 69, pp. 275–294, 1998.

[103] J. M. Ver Hoef, N. Cressie, and R. P. Barry, "Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT)," *Journal of Computational and Graphical Statistics*, vol. 13, no. 2, pp. 265–282, 2004.

[104] H. Wackernagel, *Multivariate Geostatistics*. New York: Springer-Verlag Heidelberg, 2003.
[105] G. Wahba, *Spline Models for Observational Data*. SIAM, First ed., 1990.
[106] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283–298, 2008.
[107] C. K. Wikle, "A kernel-based spectral model for non-Gaussian spatio-temporal processes," *Statistical Modelling*, vol. 2, pp. 299–314, 2002.
[108] C. K. Wikle, "Hierarchical Bayesian models for predicting the spread of ecological processes," *Ecology*, vol. 84, no. 6, pp. 1382–1394, 2003.
[109] C. K. Wikle, L. M. Berliner, and N. Cressie, "Hierarchical Bayesian space-time models," *Environmental and Ecological Statistics*, vol. 5, pp. 117–154, 1998.
[110] C. K. Williams and D. Barber, "Bayesian Classification with Gaussian processes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1342–1351, 1998.
[111] I. Woodward, M. R. Lomas, and R. A. Betts, "Vegetation-climate feedbacks in a greenhouse world," *Philosophical Transactions: Biological Sciences*, vol. 353, no. 1365, pp. 29–39, 1998.
[112] Y. Xue, X. Liao, and L. Carin, "Multi-task learning for classification with Dirichlet process priors," *Journal of Machine Learning Research*, vol. 8, pp. 35–63, 2007.
[113] K. Yu, V. Tresp, and A. Schwaighofer, "Learning Gaussian processes from multiple tasks," in *Proceedings of the International Conference on Machine Learning (ICML 2005)*, pp. 1012–1019, 2005.
[114] H. Zhang, "Maximum-likelihood estimation for multivariate spatial linear coregionalization models," *Environmetrics*, vol. 18, pp. 125–139, 2007.