ComplexDepend
Methodology: Statistics for complex stochastic dependence

Innovation area: Method

Key Innovator: Arnoldo Frigessi

Partners: UiO, NR, NTNU and all other partners

Research staff scientists: Nils L. Hjort (principal investigator, UiO), Odd Aalen (principal investigator, UiO), Arnoldo Frigessi (principal investigator, UiO), Marit Holden (principal investigator, NR), Håvard Rue (principal investigator, NTNU), Kjersti Aas (principal investigator, NR), Ørnulf Borgan (principal investigator, UiO), Magne Aldrin (NR), Lars Holden (NR), Gunnhildur H. Steinbakk (NR), Arne Bang Huseby (UiO), Ingrid Glad (UiO), Knut Liestøl (UiO), Tron Anders Moger (UiO), Magne Thoresen (UiO), Henning Omre (NTNU), Birgitte F. de Blasio (principal investigator, UiO), Egil Ferkingstad (NR), Ingrid Hobæk Haff (NR), Geir Storvik (UiO), Jo Eidsvik (NTNU), Ida Scheel (UiO), Peter Guttorp (Univ. Washington and NR), Thordis Thorarinsdottir (NR) and others

Ph.D students: Gudmund H. Hermansen (UiO/NR), Linn Cecilie Bergersen (UiO), Kristoffer Hellton (UiO), Øystein Sørensen (UiO), and others

Postdocs: Kjetil Røysland (UiO), Daniel Simpson (NTNU), Kukatharmini Tharmaratnam (UiO), Geir Kjetil Sandve (UiO), Alex Lenkoski (NR) and others

International contacts and collaborators: Sylvia Richardson (Imperial College, London), Henry Wynn (LSE, London), Paul Embrechts (ETH Zürich), Stephen Walker (Kent, UK), Åke Svensson (Stockholm), Tom Britton (Stockholm), Gianpaolo Scalia Tomba (University of Rome II, Italy), Antonella Villani (L'Aquila), Ismail Ahmed (Inserm), Elja Arjas (Helsinki University, Finland) and more (see paper co-authors)

Scope:

Produce scientific results which underpin (sfi)²'s innovation strategy by providing new statistical methods and computational tools to current and future innovation projects. Progress statistical science in general.

Results in 2012 and plans:

The mission of (sfi)² is to develop core statistical methodologies, strategically necessary to achieve innovation goals in four key sectors. By carefully selecting methodological research themes, following our understanding of the innovation challenges in the four areas and at our partners, insisting on our current methodological strengths, we shall produce basic statistical science that will lead to competitive advantages. We maintain a fertile scientific ground to grow the new generations of high quality statisticians. Producing new statistical methods which attack these issues enable us to develop innovative approaches which

(a) explain behaviours and tests hypothesis,
(b) predict future states, and
(c) find causal relations,
(d) with uncertainty.

In all these research directions we have top international competence.

This is our core methodology project, where many generic statistical methods are developed, under the influence of our projects, and feeding back into our projects. Here are themes where we expect to publish in 2013:

p >> n regression models in prediction and variable selection, under sparsity assumptions. Particular attention to modelling of interactions. This leads to ultra-high dimension p that requires preselection. We will develop safe preselections methods. A second direction will be to apply lasso to non-linear monotone models.

Model comparison, in order to select adequate models for the actual inferential goals. Here we will look to various models of sparsity and consider selection mechanisms.

Statistical causality, in order to understand cause-effect dynamics from data.

Semi-automatic Monte Carlo based inference and inla algorithms. Extensions of inla beyond the Gaussian priors are investigated.

Pair copula constructions, to capture co-movements in the tails. We want to investigate the connection to Bayesian network estimation. Focused truncation techniques will be investigated. An application in climate science is underway.

Design of experiments for Bayesian Networks.

Data integration, which incorporates functional mechanism that order the various data sources.

Direct Bayesian estimation in Gaussian Graphical Models. We will investigate methods for estimating sparse inverse covariance matrices using a new class of prior distributions. Recent breakthroughs have revealed the potential for direct Bayesian estimation in this class of models, where computationally intensive approximate methods were previously necessary. We will continue this research, which will dramatically improve the scope of sparse methods in high-dimensional hierarchical modelling.

Functional vector autoregression through manifold embedding and array-variate graphical models. We will investigate Bayesian methods for estimating sets of correlated random functions. The dimensionality of such a problem requires functions to be embedded in a lower dimensional manifold, thereby sparsely representing a collection of functions as an array of data. By then modelling the elements of this array sparsely, methods related to Gaussian graphical models (see point 1 above) can be utilized to quickly perform inference. We will develop this methodology, which requires creating a new class of prior distributions for precision matrices with diagonal constraints.

Bayesian Model Averaging in Endogenously Determined Systems of Equations. Many situations in economics and medicine encounter the difficulty that the covariates are correlated with the residuals of the dependent variable. This endogenous variable framework requires special attention to estimate causal effects correctly. Incorporating variable uncertainty has previously been handled poorly in these systems. We will develop a methodology that enables Bayesian model averaging to be performed quickly and easily in the endogenous variable model.

We will focus on the training profile of our students and postdocs, support their first experiences with research work and the cultural environment.

Papers:

Aalen, OO. Armitage lecture 2010: Understanding treatment effects: the value of integrating longitudinal data and survival analysis. Statistics in Medicine, 2012; Vol. 31: 1903–1917.

Aalen, OO. Survival analysis, longitudinal analysis and causal inference. Lifetime Data Analysis, 2010; Vol. 16(1): 1-2.

Aalen, OO. Understanding disease processes. Statistics in Medicine, 2010; Vol. 29(11): 1159-1160.

Aalen, Odd O.; Gunnes, Nina. A dynamic approach for reconstructing missing longitudinal data using the linear increments model. Biostatistics, 2010; Vol. 11(3): 453-472.

Aalen, OO; Røysland, K; Gran, JM; Ledergerber, B. Causality, mediation and time: a dynamic viewpoint. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2012; Vol. 175(4): 831–861.

Aas, Kjersti; Czado, Claudia; Frigessi, Arnoldo; Bakken, Henrik: Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics, 2009; Vol. 44 (2): 182-198.

Aldrin, Magne; Hobæk Haff, Ingrid; Rosland, Pål: The effect of salting with magnesium chloride on the concentration of particular matter in a road tunnel. Atmospheric Environment, 2008; Vol. 42: 1762-1776.

Aune, E; Eidsvik, J; Pokern, Y. Iterative numerical methods for sampling from high dimensional Gaussian distributions. Statistics and Computing, 2012; doi=10.1007/s11222-012-9326-8.

Baghishani, H; Rue, H; Mohammadzadeh, M. On a hybrid data cloning method and its application in generalized linear mixed models. Statistics and Computing, 2012; Vol. 22(2): 597–613.

Beguin, J; Martino, S; Rue, H; Cumming, SG. Hierarchical analysis of spatially autocorrelated ecological data using integrated nested Laplace approximation. Methods in Ecology and Evolution, 2012; Vol. 3: 921–929.

Bergersen, Linn Cecilie; Glad, Ingrid Kristine; Lyng, Heidi: Weighted Lasso with Data Integration. Statistical Applications in Genetics and Molecular Biology, 2011; Vol. 10(1).

Bolin, David; Lindgren, Finn: Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. Annals of Applied Statistics, 2011; Vol. 5(1): 523-550.

Bray F, Haugen M, Moger TA, Tretli S, Aalen OO, Grotmol T: Age-incidence curves of nasopharyngeal carcinoma worldwide: bimodality in low-risk populations and aetiologic implications. Cancer Epidemiol Biomarkers Prev., 2008; Vol. 17(9): 2356-2365.

Brechmann, Eike C; Czado, Claudia; Aas, Kjersti: Truncated regular vines in high dimensions with application to financial data. Canadian Journal of Statistics, 2012; Vol. 40(1): 68-85.

Brown, Andrew Anand; Richardson, Sylvia; and Whittaker, John: Application of the Lasso to Expression Quantitative Trait Loci Mapping. Statistical Applications in Genetics and Molecular Biology, 2011; Vol. 10(1).

Buonaccorsi, JP; Laake, P; Veierød, MB. On the power of the Cochran-Armitage test for trend in the presence of misclassification. Statistical Methods in Medical Research, 2011, doi: 10.1177/0962280211406424.

Bøvelstad, Hege M.; Borgan, Ørnulf: Assessment of evaluation criteria for survival prediction from genomic data. Biometrical Journal, 2011; Vol. 53(2): 202-216.

Cameletti, M; Lindgren, F; Simpson, D; Rue, H. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. AStA Advances in Statistical Analysis, 2012; doi: 10.1007/s10182-012-0196-3.

Cheng, Y; Lenkoski, A. Hierarchical Gaussian graphical models: Beyond reversible jump. Electronic Journal of Statistics, 2012; Vol. 6: 2309-2331.

Claeskens, G; Hjort, NL. Minimizing average risk in regression models. Econometric Theory, 2008; Vol. 24: 493-527.

De Blasi P, Hjort N. L: The Bernstein–von Mises theorem in semiparametric competing risks models. Journal of Statistical Planning and Inference, 2009; Vol. 139(7): 2316-2328.

De Blasio, Birgitte Freiesleben; Seierstad, Taral Guldahl; Aalen, Odd O.: Frailty effects in networks: comparison and identification of individual heterogeneity versus preferential attachment in evolving networks. Journal of the Royal Statistical Society, Series C: Applied Statistics, 2011; Vol. 60(2): 239- 259.

Eidsvik, Jo; Finley, Andrew O.; Banerjee, Sudipto; Rue, Håvard: Approximate Bayesian inference for large spatial datasets using predictive process models. Computational Statistics & Data Analysis, 2012; Vol. 56(6): 1362-1380.

Eidsvik, J; Martino, S; Rue, H. Approximate Bayesian inference in spatial generalized linear mixed models. Scandinavian Journal of Statistics, 2009; Vol. 36(1): 1-22.

Fong, Youyi; Rue, Håvard; Wakefield, Jon. Bayesian inference for generalized linear mixed models. Biostatistics, 2010; Vol. 11(3): 397-412.

Frigessi, Arnoldo; Løland, Anders; Pievatolo, Antonio; Ruggeri, Fabrizio: Statistical rehabilitation of improper correlation matrices. Quantitative Finance, 2011; Vol. 11(7): 1081-1090.

Fruwirth-Schnatter, S; Fruwirth, R; Held, L; Rue, H. Improved auxiliary mixture sampling for hierarchical models of non-gaussian data. Statistics and Computing, 2009; Vol. 19(4): 479-492.

Frøslie, KF; Røislien, J; Laake, P; Henriksen, T; Qvigstad, E; Veierød, MB. Categorisation of continuous exposure variables revisited. A response to the Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study. BMC Medical Research Methodology, 2010 ; Vol. 10:103.

Gandy A, Therneau TM, Aalen OO: Global tests in the additive hazards regression model. Statistics in Medicine, 2008; Vol. 27(6): 831-844.

Gjessing, Håkon K.; Røysland, Kjetil; Aalen, Odd O.; Pena, Edsel A. Recurrent events and the exploding Cox model. Lifetime Data Analysis, 2010; Vol. 16(4): 525-546.

Grønneberg, Steffen; Hjort, Nils Lid: On the errors committed by sequences of estimator functionals. Mathematical Methods of Statistics, 2011; Vol. 20(4): 327- 346.

Gunnes N, Seierstad TG, Aamdal S, Brunsvig PF, Jacobsen AB, Sundstrøm S, Aalen OO. Assessing quality of life in a randomized clinical trial: correcting for missing data. BMC Med Res Methodol., 2009; Vol. 9:28.

Guttorp, Peter: The role of statisticians in international science policy. Environmetrics, 2011; Vol. 22(7): 817-825.

Guttorp, P; Thorarinsdottir, TL. Bayesian Inference for Non-Markovian Point Processes. In: Advances and Challenges in Space-time Modelling of Natural Events, Lecture Notes in Statistics 2012; Vol. 207: 79-102.

Hammer, Hugo Lewi; Tjelmeland, Håkon: Control variates for the Metropolis-Hastings algorithm. Scandinavian Journal of Statistics, 2008; Vol. 35: 400-414.

Hammer, Hugo; Tjelmeland, Håkon: Approximate forward–backward algorithm for a switching linear Gaussian model. Computational Statistics & Data Analysis, 2011; Vol. 55(1): 154-167.

Hansen, LV; Thorarinsdottir, TL. A note on moving average models for Gaussian random fields. Statistics & Probability Letters, 2013; Vol. 83(3): 850-855.

Hjort, NL. Approximating Data. Journal of the Korean Statistical Society, 2008; Vol. 37/3,: 221-225.

Hjort, NL; McKeague, IW; Van Keilegom, I. Extending the Scope of Empirical Likelihood. Annals of Statistics, 2009; Vol. 37(3): 1079-1111.

Hjort NL, Walker SG: Quantile pyramids for Bayesian nonparametrics. Annals of Statistics, 2009; Vol. 37(1): 105-131.

Hobæk Haff I., Aas K., Frigessi A: On the simplified pair-copula construction -- simply useful or too simplistic?, J. Multiv Anal, 2010; Vol. 101(5): 1296-1310.

Holden L, Hauge R, Holden M: Adaptive independent Metropolis–Hastings. Annals of Applied Probability, 2009; Vol. 19(1): 395-413.

Laake, Ida; Thune, Inger; Selmer, Randi Marie; Tretli, Steinar; Slattery, Martha L; Veierød, Marit B: A prospective study of body mass index, weight change, and risk of cancer in the proximal and distal colon. Cancer Epidemiology, Biomarkers and Prevention, 2010; Vol. 19(6): 1511-1522.

Lawyer G, Ferkingstad E, Nesvåg R, Varnäs K, Agartz I: Local and covariate-modulated false discovery rates applied in neuroimaging. NeuroImage, 2009; Vol. 47(1): 213-219.

Li, Y; Brown, P; Rue, H; al-Maini, M; Fortin, P. Spatial modelling of lupus incidence over 40 years with changes in census areas. Journal of the Royal Statistical Society: Series C (Applied Statistics), 2012; Vol. 61: 99–115.

Lindgren F., Lindstrøm, J. and Rue, H: An explicit link between Gaussian fields and Gaussian Markov random fields: The SPDE approach, JRSS(B) discussion paper, 2011.

Lindgren, Finn; Martins, Thiago; Rue, Håvard; Simpson, Daniel: Discussion on "Spatial prediction in the presence of positional error". Environmetrics, 2011; Vol. 22(2): 127.

Lindgren, Finn; Rue, Håvard; Lindström, Johan: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 2011; Vol. 73(4): 423-498.

Lindgren, Georg; Lindgren, Finn: Stochastic Asymmetry Properties of 3D Gauss-Lagrange Ocean Waves with Directional Spreading. Stochastic Models, 2011; Vol. 27(3): 490-520.

Lydersen S, Fagerland MW, Laake P: Recommended tests for association in 2x2 tables. Statistics in Medicine, 2009; Vol. 28(7): 1159-1175.

Martinelli, G; Eidsvik, J; Hauge, R; Førland, MD. Bayesian networks for prospect analysis in the North Sea. The American Association of Petroleum Geologists, 2011; Vol. 95(8): 1423-1442.

Martino, S; Akerkar, R; Rue, H. Approximate Bayesian Inference for Survival Models. Scandinavian Journal of Statistics, 2011; Vol. 38(3): 514-528.

Martino S, Rue H: Case Studies in Bayesian Computation using INLA in "Complex data modeling and computationally intensive statistical methods" Secchi P. and Mantovan P. (editors), 2010.

Martinussen T, Aalen OO, Scheike TH: The Mizon-Richard encompassing test for the Cox and Aalen additive hazards models. Biometrics, 2008; Vol. 64(1): 164-171.

Moger TA, Aalen OO: Regression models for infant mortality data in Norwegian siblings, using a compound Poisson frailty distribution with random scale. Biostatistics, 2008; Vol. 9(3): 577-591.

Moger, TA; Haugen, M; Yip, BHK; Gjessing, HK; Borgan, Ø. A hierarchical frailty model applied to two-generation melanoma data. Lifetime Data Analysis, 2011; Vol. 17(3): 445-460.

Moger TA, Pawitan Y, Borgan Ø: Case-cohort methods for survival data on families from routine registers. Statistics in Medicine, 2008; Vol. 27(7): 1062-1074.

Möller, A; Lenkoski, A; Thorarinsdottir, TL. Multivariate probabilistic forecasting using Bayesian model averaging and copulas. In press at Quarterly Journal of the Royal Meteorological Society, 2012.

Nygård, S; Borgan, Ø; Lingjærde, OC; Størvold, HL. Partial least squares Cox regression for genome-wide data. Lifetime Data Analysis, 2008; Vol. 14(2): 179-195.

Parr CL, Hjartåker A, Scheel I, Lund E, Laake P, Veierød MB: Comparing methods for handling missing values in food-frequency questionnaires and proposing k nearest neighbours imputation: effects on dietary intake in the Norwegian Women and Cancer study (NOWAC). Public Health Nutr., 2008; Vol. 11(4): 361-70.

Paul, M; Riebler, A; Bachmann, LM; Rue, Håvard; Held, L. Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. Statistics in Medicine, 2010; Vol. 29(12): 1325-1339.

Rezaie, J; Eidsvik, J. Shrinked ($1+\alpha$) ensemble Kalman filter and α Gaussian mixture filter. Computational Geosciences, 2012; Vol. 16(3): 837-852.

Riebler, A; Held, L; Rue, H. Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations. Annals of Applied Statistics, 2012; Vol. 6(1): 304-333.

Riebler, A; Held, L; Rue, H; Bopp, M. Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2012; Vol. 175: 473–490.

Rue H., Martino S., Chopin N: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Statist. Soc. B, 2009; Vol. 71(2): 319-392.

Røysland, Kjetil: A martingale approach to continuous-time marginal structural models. Bernoulli, 2011; Vol. 17(3): 895-915.

Samuelsen SO, Eide GE: Attributable fractions with survival data. Statistics in Medicine, 2008; Vol. 27(9): 1447-67.

Sandve, Geir Kjetil; Gundersen, Sveinung; Rydbeck, Halfdan; Glad, Ingrid Kristine; Holden, Lars; Holden, Marit; Liestøl, Knut; Clancy, Trevor; Ferkingstad, Egil; Johansen, Morten; Nygaard, Vegard; Tøstesen, Eivind; Frigessi, Arnoldo; Hovig, Eivind: The Genomic HyperBrowser: inferential genomics at the sequence level. Genome Biology, 2010; Vol. 11(12): R121.

Scheel, I; Green, PJ; Rougier, JC. A Graphical Diagnostic for Identifying Influential Model Choices in Bayesian Hierarchical Models. Scandinavian Journal of Statistics, 2011; Vol. 38(3): 529-550.

Scheel, I; Hinnerichsen, M. The Impact of Climate Change on Precipitation-related Insurance Risk: A Study of the Effect of Future Scenarios on Residential Buildings in Norway. The Geneva Papers, 2012; Vol. 37: 365–376.

Schmidt, AM; Guttorp, Peter; O'Hagan, A.: Considering covariates in the covariance structure of spatial processes. Environmetrics, 2011; Vol. 22(4): 487-500.

Sexton J, Laake P: Standard errors for bagged and random forest estimators. CSDA, 2009; Vol. 53: 801-811.

Sexton J, Laake P: Stochastic Approximation Boosting for Incomplete Data Problems. Biometrics, 2009; Vol. 65(4): 1156–1163.

Simpson, Daniel; Illian, Janine Baerbel; Lindgren, Finn; Sørbye, Sigrunn; Rue, Håvard: Going off grid: Computationally efficient inference for log-Gaussian Cox processes. Journal of the Royal Statistical Society Series B: Statistical Methodolgy, to appear in 2012.

Steinbakk, Gunnhildur Högnadóttir; Storvik, Geir: Posterior predictive p-values in Bayesian Hierarchical models. Scandinavian Journal of Statistics, 2009; Vol. 36(2): 320-336.

Storvik, Bård; Storvik, Geir, Fjørtoft, Roger. On the combination of correlated images using meta-Gaussian distributions. IEEE Transactions on Geoscience and Remote Sensing, 2009; Vol. 47(7): 2372-2379.

Storvik, Geir: On the flexibility of Metropolis-Hastings acceptance probabilities in auxiliary variable proposal generation. Scandinavian Journal of Statistics, 2011; Vol. 38(2): 342-358.

Thorarinsdottir, TL; Scheuerer, M; Feldmann, K. Statistical post-processing of ensemble forecasts. Promet, 2012; Vol. 37(3/4): 43-52.

Valberg, M; Grotmol, T; Tretli, S; Veierød, MB; Devesa, SS; Aalen, OO. Frailty modeling of age–incidence curves of osteosarcoma and Ewing sarcoma among individuals younger than 40 years. Statistics in Medicine, 2012; Vol. 31(28): 3731-3747.

Technical reports:

Hansen, LV; Thorarinsdottir, TL. A note on moving average models for Gaussian random fields. CSGB Research Report 13/2012, 2012.

Hobæk Haff, Ingrid: Parameter estimation for pair-copula constructions, NR-Note, SAMBA/36/10, 2010.

Lindgren F, Lindstrøm, J, Rue, H: An explicite link between Gaussian fields and Gaussian Markov random fields: The SPDE approach. Technical report statistics 5, Department of mathematical sciences, NTNU, 2010.

Løland, Anders; Wilhelmsen, Mathilde: Monte Carlo simulation based validation. NR-Note, SAMBA/32/08, 2008.

Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O, Hinnerichsen, M. & Meze-Hausken, E: A Bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. Derivation of distributions and MCMC sampling schemes. Statistical Research Report, No. 2, Department of Mathematics, University of Oslo, 2010.

Scheel, I., Green, P. J. & Rougier, J. C: Applications of the Local critique plot. Statistical Research Report, No. 7, Department of Mathematics, University of Oslo, 2010.

Books:

OO Aalen, Ø Borgan, HK Gjessing - Survival and Event History Analysis: A Process Point of View- 2009 – Springer.

G Claeskens, NL Hjort - Model Selection and Model Averaging, 2008 - Cambridge Univ. Press.

Inge Helland - Steps Towards a Unified Basis for Scientific Models and Methods. World Scientific, Singapore, 2010.