• Example: Alcohol related deaths in Finland
• Spatial priors and benefits of GP prior
• Computation and approximations
• Spatio-temporal
• Explanatory variables
• Integration over the latent space
• Hyperparameters
Example: deaths in Finland

(a) Number of deaths

Disease mapping with Gaussian processes
Example: deaths in Finland

(d) Number of deaths

(e) Raw relative risk

Aki.Vehtari@aalto.fi

Disease mapping with Gaussian processes
Example: deaths in Finland

(g) Number of deaths
(h) Raw relative risk
(i) Smoothed risk
Example: alcohol related diseases in Finland

- Collaboration: The National Institute for Health and Welfare
- Data: Statistics Finland
- Population of Finland: \( \approx 5.3 \) million
- About 10 500 inhabited 5km \( \times \) 5km cells in Finland
  - many cells with no inhabited neighbors
- In 2001–2005 about 7 900 died due to alcohol diseases (more than five times compared to deaths due to traffic)
  - expected death count less than one per cell
Example: alcohol related diseases in Finland

- Sex-age-education standardized expected death counts used to compute the raw risk
- Risk smoothed using GP with long and short length scale and negative-binomial observation model
Example: alcohol related diseases in Finland

- Sex-age-education standardized expected death counts used to compute the raw risk
- Risk smoothed using GP with long and short length scale and negative-binomial observation model

Is the relative risk higher in the population centers?

Aki.Vehtari@aalto.fi

Disease mapping with Gaussian processes
Example: alcohol related diseases in Finland

- The smoothed relative risks vs. the population density
Example: alcohol related diseases in Finland

- The smoothed relative risks vs. the population density

- Add population density as explanatory variable
Example: alcohol related diseases in Finland

- Population density and spatial variation explain the variation in the risk

**Population density effect**

**Spatial effect**

Disease mapping with Gaussian processes
Adding explanatory covariate can change the picture

1) Spatial

2) Spatial+covariate

In spatial epidemiology CAR is most used model.

Correlation defined conditionally based on a neighborhood structure → discrete definition

- major computational speed-up if a precision matrix is sparse due to small neighborhoods
- describes only local correlation
- neighborhood definition may be difficult for irregularly spaced data and high dimensional data
Example: alcohol related diseases in Finland
Comparison to CAR

- Compared to CAR computed with INLA software
  - CAR model lacks long range correlation part
  - CAR model has much higher variance, especially for cells having no or few inhabited neighbors
  - GP has a better predictive performance
• Markov random field prior can be good
  - e.g. INLA-software can approximate Matérn covariance function with MRF
  - but precision matrix is not going to be sparse in high dimensional cases ($d \geq 3$), e.g. INLA-software doesn’t support $d > 3$ and limited support for $d = 3$
Computation and approximations

- Full $O(n^3)$
- Short range dependencies
  - Markov $\rightarrow$ sparse precision matrix
  - Compact support $\rightarrow$ sparse covariance matrix
  - $O(p^3 n^3)$, where $0 < p < 1$ is the proportion of non-zeros
- Long range dependencies
  - Reduced rank (e.g. FIC) $O(nm^2)$
  - SVI-GP $O(m^3)$ (Hensman et al, 2013)
The correlation structure of FIC with different choices of inducing inputs

Figure: The correlation for 3 locations $\mathbf{x}$. Inducing inputs are marked with $\star$. 

Reduced rank approximations and inducing points
• No single approximation which works efficiently for both short and long range dependencies

- e.g. compact support + FIC (used in alcohol study)

• No single approximation which works efficiently for both short and long range dependencies

• Short and long range dependencies
  - e.g. compact support + FIC (used in alcohol study)
  Vanhatalo, Pietiläinen, Vehtari, Stat in med, 2010,
  http://dx.doi.org/10.1002/sim.3895
Spatio-temporal

- Full $O(n^3 T^3)$
- Markov / compact support / reduced rank
- INLA-software: unstructured interaction (ie. no model for spatio-temporal jointly)
- Cseke et al - discrete spatio-temporal model, sparse precision, restricted sparse messages
- infinite-dimensional filtering $O(n^3 T) \ (O(nm^2 T))$
  Simo Särkkä talks about this tomorrow
County incidences and background population for years 1953–2003.

51 years, 431 counties → 21,981 observations

Data: Finnish Cancer Registry

Model: GP with temporal + spatial + spatiotemporal component
Example: lung cancer women

(a) Temporal

(b) Spatial

Disease mapping with Gaussian processes
Example: lung cancer women

(a) 1953
(b) 1963
(c) 1973
(d) 1983
(e) 1993
(f) 2003

Disease mapping with Gaussian processes
Spatio-temporal

- Spatio-temporal GPs can be written as linear stochastic partial differential equations (SPDE)
- Reduces computational complexity from $O(n^3 T^3)$ to $O(n^3 T)$, i.e. method scales linearily in $T$
- SPDEs make it easier to specify non-stationary temporal dynamics, which are necessary, for example, when performing future predictions
- $n$ limited as for spatial GP
  - few thousand with no sparse approximations
  - more than ten thousand with sparse approximations
- Has been tested with over million spatio-temporal points
- Simo Särkkä talks more about this tomorrow
Spatio-temporal malaria models?

- Spatio-temporal GPs can be written as linear stochastic partial differential equations (SPDE).
- SPDEs make it easier to specify non-stationary temporal dynamics, which are necessary, for example, when performing future predictions:
  - seasonal variation
  - transmission dynamics with SPDEs?
Non-stationarity

- SPDEs make it easier to specify non-stationary temporal dynamics
- Spatial non-stationarity
  - deformations
  - additional GP for latent signal magnitude or length-scale
Explanatory covariates

- Goal is to explain the spatial variation
- Spatial maps can be used to aid hypothesis generation
- Adding covariates hopefully makes the residual in spatial domain unstructured
- GP can model non-linearities and interactions implicitly
1043 cases of acute myeloid leukemia in adults
- recorded between 1982 and 1998 in the North West Leukemia Register in the United Kingdom
- log-logistic model for survival times (16% were censored)
- predictors are
  - age
  - sex
  - white blood cell count (WBC) at diagnosis
  - the Townsend score which is a measure of deprivation for district of residence
Figure: Posterior mean of the latent function
Leukemia survival times

- **Age (years)**
- **Sex**
- **WBC (log$_{10}(50\times10^9/L)$)**
- **Townsend deprivation index (TDI)**
- **Spatial location**

Disease mapping with Gaussian processes

Aki.Vehtari@aalto.fi
Leukemia survival times

Age (years) vs. Survival

Sex vs. Survival

WBC (log$_{10}$($50 \times 10^9$/L)) vs. Survival

Townsend deprivation index (TDI) vs. Survival

Spatial location with color scale

Aki.Vehtari@aalto.fi

Disease mapping with Gaussian processes
Leukemia survival times

Aki.Vehtari@aalto.fi

Disease mapping with Gaussian processes
Leukemia survival times

Analysis in GP chapter of

Explanatory covariates

- GP can model non-linearities and interactions implicitly
- INLA-software using MRFs allows additive effects and 2D interactions
Multiple diseases

- Multitask / multioutput GPs
  - just add the disease type as a covariate
Integration over the latent space

- Non-Gaussian models, e.g., \( y \sim \text{Poisson}(\alpha \exp(f(s, \theta))) \)
- We are interested in predictions \( p(y_i|s_i) \)
- Integration over the latent variables \( f_i \) and hyperparameters \( \theta \) required
In our experiments
- EP about as good as MCMC, but **much** faster
- Laplace almost as good as EP, but somewhat faster
- VB not as good as EP, byt YMMV
- difference is negligible for many likelihoods given larger datasets
- differences in classification and with non-log-concave likelihoods
Integration over the latent space

- In our experiments
  - EP about as good as MCMC, but much faster
  - Laplace almost as good as EP, but somewhat faster
  - VB not as good as EP, by YMMV
  - difference is negligible for many likelihoods given larger datasets
  - differences in classification and with non-log-concave likelihoods

- Mysterious Sheffield-method? (Hensman et al, submitted)
Integration over the latent space

- In our experiments
  - EP about as good as MCMC, but much faster
  - Laplace almost as good as EP, but somewhat faster
  - VB not as good as EP, byt YMMV
  - difference is negligible for many likelihoods given larger datasets
  - differences in classification and with non-log-concave likelihoods

- Mysterious Sheffield-method? (Hensman et al, submitted)

- I think that in most cases distributional approximations ok
  - If not, pseudo-marginal likelihood approach (Filippone & Girolami, 2013) might be the best choice for MCMC
Hyperparameter inference

- Type II MAP
  - works well when the number of hyperparameters is small and $n$ is big
- Adaptive grid 1–3 hyperparameters
- CCD
  - 1–15 hyperparameters $\rightarrow$ 3–287 integration points
  - usually works well, but sometimes underestimates the uncertainty

Aki.Vehtari@aalto.fi

Disease mapping with Gaussian processes
Hyperparameter inference

- Type II MAP
  - works well when the number of hyperparameters is small and $n$ is big
- Adaptive grid 1–3 hyperparameters
- CCD
  - 1–15 hyperparameters $\rightarrow$ 3–287 integration points
  - usually works well, but sometimes underestimates the uncertainty
- Linear approximation (Garnett, Osborne, Hennig, 2013)
- EP can be used to integrate over noise and signal variances (other hyperparameters in theory, but not fast (yet?))
- MCMC
Hyperparameters

(a) Grid based

(b) Monte Carlo

Figure: The grid based, Monte Carlo and central composite design integration. Contours show the posterior density $q(\log(\vartheta) | D)$ and the integration points are marked with dots.

Disease mapping with Gaussian processes

Aki.Vehtari@aalto.fi
Code available in Matlab/Octave (RccpOctave for R) toolbox GPstuff


GPstuff homepage: http://becs.aalto.fi/en/research/bayes/gpstuff/
GPstuff

• Sparse models
  - Compactly supported covariance functions
  - Fully and partially independent conditional (FIC, PIC)
  - Compactly supported plus FIC (CS+FIC)
  - Variational sparse (VAR), Deterministic training conditional (DTC), Subset of regressors (SOR)

• Latent inference
  - marginal posterior corrections (cm2 and fact)
  - Scaled Metropolis, Scaled HMC, Elliptical slice sampling

• Hyperparameter inference
  - Type II ML/MAP
  - Leave-one-out cross-validation (LOO-CV)
  - Metropolis, HMC, No-U-Turn-Sampler (NUTS), Slice Sampling (SLS), Surrogate SLS, Shrinking-rank and Cov-matching SLS
  - Grid, CCD, Importance sampling
Acknowledgments

- Researchers
  - Jaakko Riihimäki
  - Ville Tolvanen
  - Simo Särkkä
  - Jarno Vanhatalo
  - Jouni Hartikainen
  - Ville Pietiläinen

- Collaborators and data
  - Heikki Joensuu, MD
  - Helsinki University Central Hospital
  - The National Institute for Health and Welfare
  - Finnish Cancer Registry
  - Statistics Finland