Bayesian covariate models in extreme value analysis

David Randell, **Philip Jonathan**, **Kathryn Turnbull**, **Mathew Jones**
EVA 2015 Ann Arbor

Acknowledgement

- Kevin Ewans, Graham Feld and other Shell colleagues.
- Colleagues in academia, especially at Lancaster University.

Hurricane Katrina

Hurricane Katrina

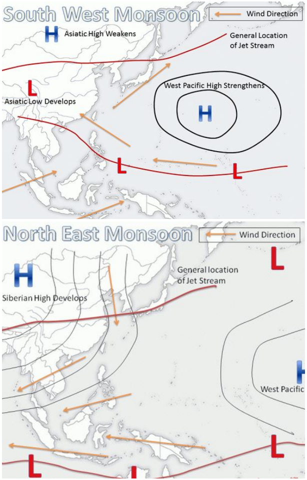
Motivation: extremes in met-ocean

- **Rational** and **consistent** design an assessment of **marine structures**
  - Reduce bias and uncertainty in estimation of **return values**.
- Non-stationary **marginal** and **conditional** extremes
  - Multiple locations, multiple variables, time-series,
  - **Multidimensional** covariates.
- Improved **understanding** and **communication** of risk
  - Incorporation within **well-established** engineering design practices,
  - **"Knock-on" effects** of "improved" inference,
  - New and existing structures.

Marginal directional-seasonal extremes

Marginal directional-seasonal extremes

- Marginal model: **single** location.
- Response: **storm peak significant wave height,** H_S^{sp}.
- Wave climate: **monsoonal**.
- Southwest monsoon ($\sim$ August, to northwest).
- Northeast monsoon ($\sim$ January, to east-northeast).
- Long **fetches** to Makassar Strait, Java Sea.
- Land shadows of Borneo (northwest), Sulawesi (northeast), Java (south).



www.westernpacificweather.com

Directional and seasonal variability

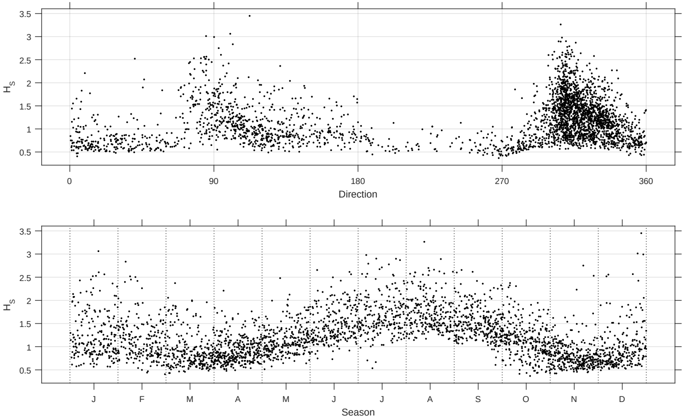


Figure: Hindcast storm peak significant wave height H_S^{sp} for 1956 – 2012 (black) on direction θ (upper panel, to which waves propogate) and season ϕ (lower panel). Also shown is sea-state significant wave height H_S (grey) on direction θ (upper panel) and season ϕ (lower panel). Northeast monsoon: August to northwest (315). Southwest monsoon: January to east-northeast (110).

Storm model

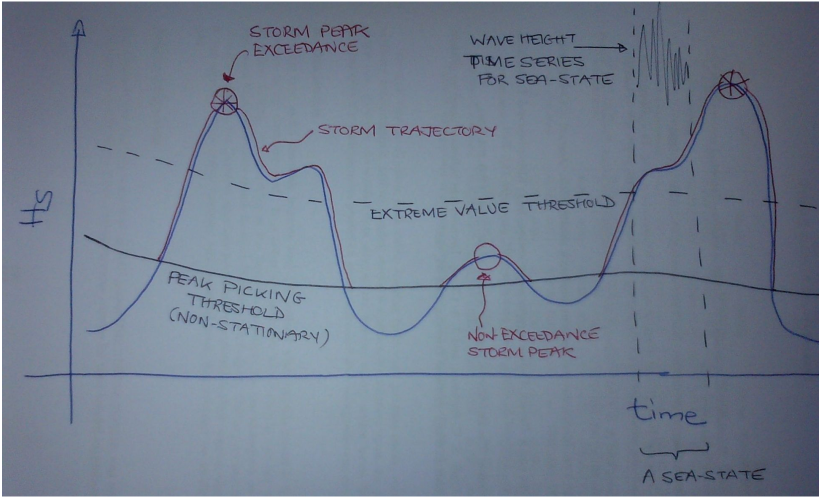


Figure: $H_S \approx 4\times$ standard deviation of ocean surface profile at a location corresponding to a specified period (typically three hours)

Model components

- Linear wave theory suggests ocean waves Rayleigh distributed, Longuet-Higgins [1952].
- Response y not exceeding the extreme value threshold ψ follows a truncated Weibull distribution with density similar to Frigessi et al. [2002]

$$f(y|\xi,\sigma,\alpha,\gamma,\psi,\tau) = \begin{cases} \tau \times f_{TW}(y|\alpha,\gamma) & \text{for } y \leq \psi \\ (1-\tau) \times f_{GP}(y|\xi,\sigma) & \text{for } y > \psi \end{cases}$$

- ψ is defined by τ, α and γ

$$\psi|\tau,\alpha,\gamma = \alpha\left(-\log(1-\tau)\right)^{\gamma} \ .$$

- No imposition of continuity in the density.
- Rate of occurrence $\rho|r$ is modelled as a Poisson distribution as in Chavez-Demoulin and Davison [2005], where r is observation counts in covariate bins.

Bayesian P-Splines

- Physical considerations suggest model parameters $\alpha, \gamma, \rho, \xi$ and σ vary smoothly with covariates θ, ϕ
- Values of $(\eta =)\alpha, \gamma, \rho, \xi$ and σ all take the form

$$\eta = B\beta_\eta$$

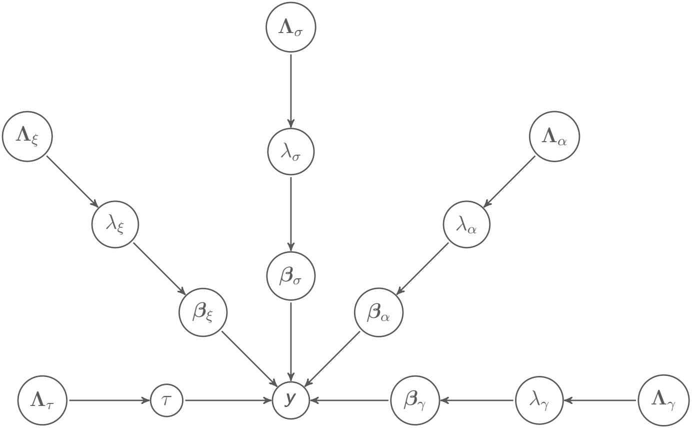
Priors:

$$\beta_\eta \sim \lambda \exp\left(-\frac{1}{2}\beta_\eta' D' D \beta_\eta\right)$$
$$\lambda \sim \mathrm{Gamma}(\mathbf{\Lambda}_\eta)$$
$$\tau \sim \mathrm{Beta}(\mathbf{\Lambda}_\tau)$$

- B is a spline basis and D is a difference matrix of order k.
- Smoothnesses λ can be estimated very easily using Gibbs sampling see Brezger and Lang [2006].
- $\mathbf{\Lambda}_\eta$ are smoothness hyper-parameters giving diffuse prior.

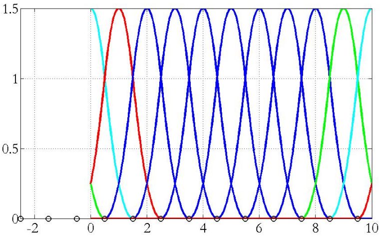
DAG for Poisson Rate of Occurrence

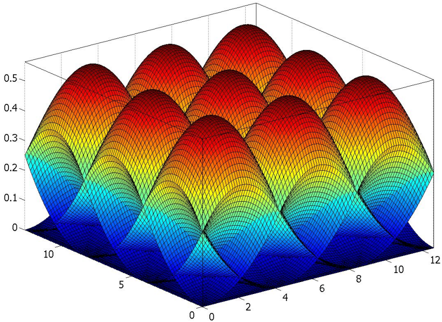
DAG for Weibull GP



Penalised B-splines

- **Wrapped** bases for periodic covariates (direction, season).
- **Multidimensional** bases easily constructed using tensor products, Eilers and Marx [2010].
- **GLAMs**, Currie et al. [2006] for efficient computation in high dimensions.





Sampling for P-Splines

- Cannot write full conditionals for generalised Pareto likelihood, so no Gibbs sampling; simple Metropolis Hastings methods don't mix well.
- Neighbouring spline parameters are highly correlated due to smoothness prior.
- Correlated spline proposals made from

$$\beta^* \sim N(\beta, G^{-1})$$

  where

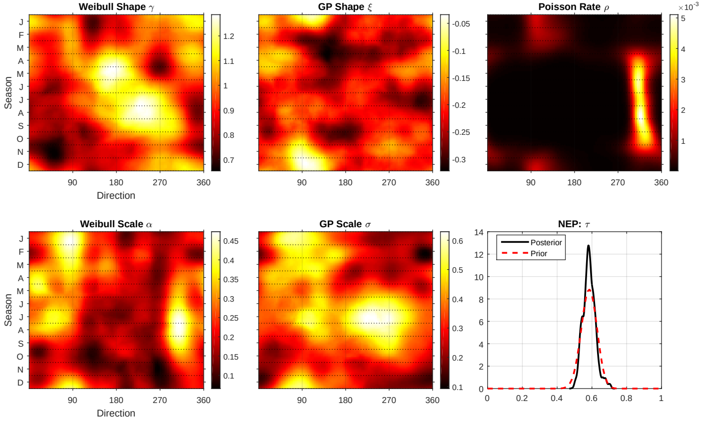
$$G = B'B + \lambda D'D \ .$$

- Gradient based MCMC methods also help to improve mixing.

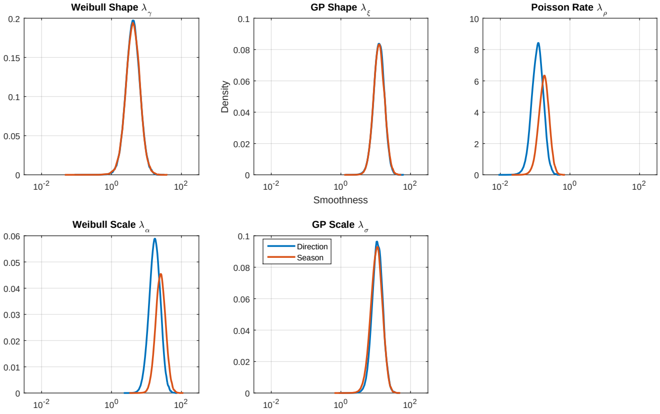
MCMC proposals exploiting gradient

- Simple Metropolis Hastings sampling is a “stochastic analogue” of penalised likelihood optimisation, but does not exploit gradient information.
- MALA and mMALA use gradient information for MCMC proposal generation. These are stochastic analogues of back-fitting and IRLS, see Roberts and Stramer [2002]. They provide random walks *downhill*, particularly important with high numbers of correlated parameters.

Parameter plots



Smoothness

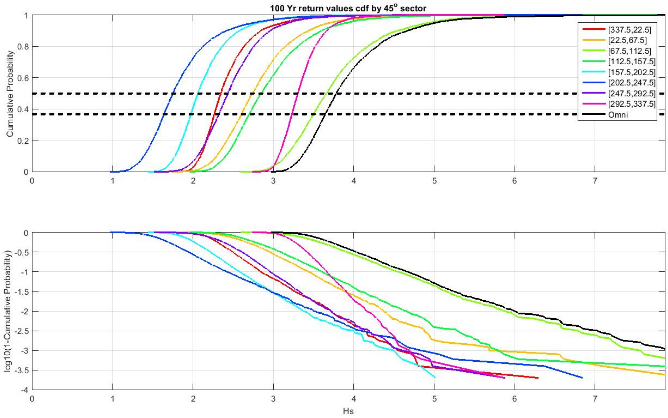


Weibull Shape λ_γ
GP Shape λ_ξ
Poisson Rate λ_ρ
Weibull Scale λ_α
GP Scale λ_σ
Density
Smoothness
Direction
Season

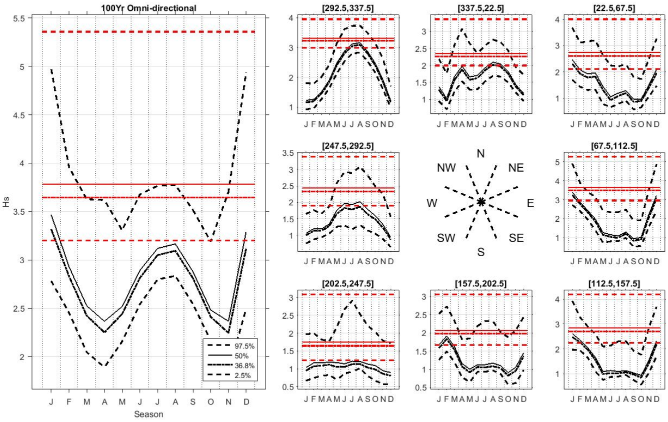
Return values

- Estimated by simulation from sample of posterior.
- H_{S100} is the maximum value of H_S^{sp} in a simulation period of 100–years.

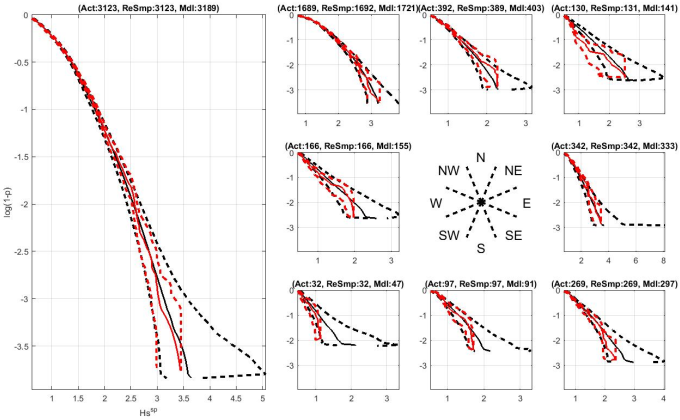
Directional 100yr return values



Directional seasonal 100yr return values



Directional Seasonal 100Yr Return Values



Summary

- Modelling non-stationarity essential for understanding extremes and development of design conditions.
- Non-parametric covariate models flexible, but need to estimate roughness.
- P-splines simple to implement and extend to periodic and higher dimensional domains.
- Bayesian P-spline for extremes
  - Roughness estimated using Gibbs sampling,
  - Different roughness for each covariate dimension,
  - Correlated MCMC proposals exploiting gradient,
  - Computationally efficient, and generally more stable than optimisation to point solution.

References

Andreas Brezger and Stefan Lang. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, 50(October 2004):967–991, 2006.

V. Chavez-Demoulin and A.C. Davison. Generalized additive modelling of sample extremes. *J. Roy. Statist. Soc. Series C: Applied Statistics*, 54:207, 2005.

I. D. Currie, M. Durban, and P. H. C. Eilers. Generalized linear array models with applications to multidimensional smoothing. *J. Roy. Statist. Soc. B*, 68:259–280, 2006.

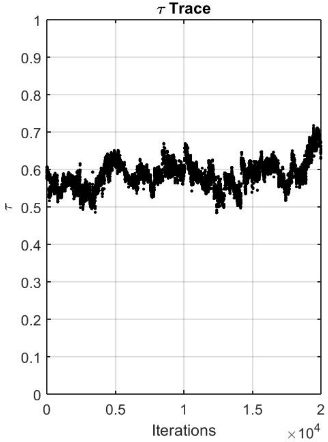
P H C Eilers and B D Marx. Splines, knots and penalties. *Wiley Interscience Reviews: Computational Statistics*, 2: 637–653, 2010.

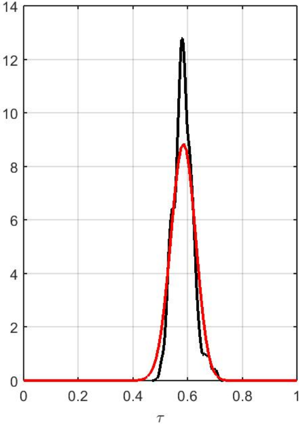
Arnoldo Frigessi, Ola Haug, and Hå vard Rue. A Dynamic Mixture Model for Unsupervised Tail Estimation without Threshold Selection. *Extremes*, 5:219–235, 2002. ISSN 1572-915X. doi: 10.1023/A:1024072610684. URL `http://link.springer.com.libproxy1.nus.edu.sg/article/10.1023/A%3A1024072610684`.

M. S. Longuet-Higgins. On the statistical distribution of the height of sea waves. *J. Mar. Res.*, 11:245–266, 1952.

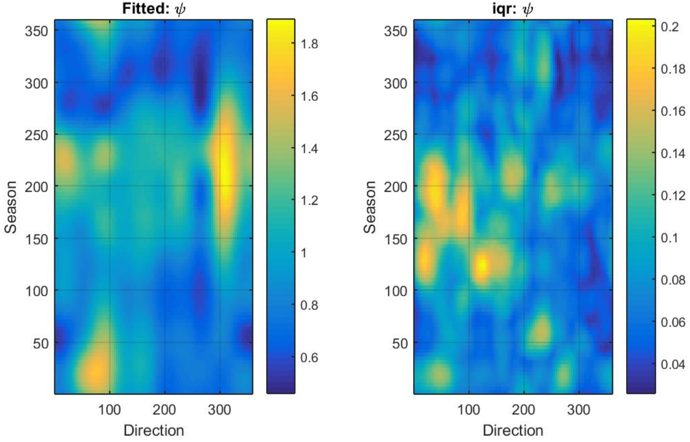
G. O. Roberts and O. Stramer. Langevin Diffusions and Metropolis-Hastings Algorithms. pages 337–357, 2002. ISSN 13875841. doi: 10.1023/A:1023562417138. URL `http://www.springerlink.com/content/nv57r5gq18088257/?p=6129beac0358495db95e120f3372c576&pi=1`.

Parameters τ

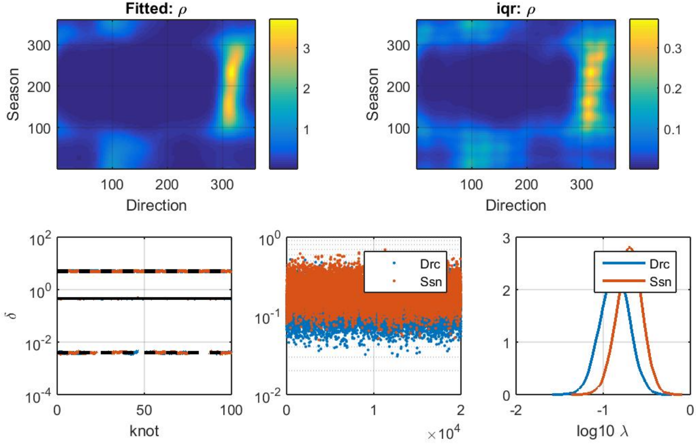




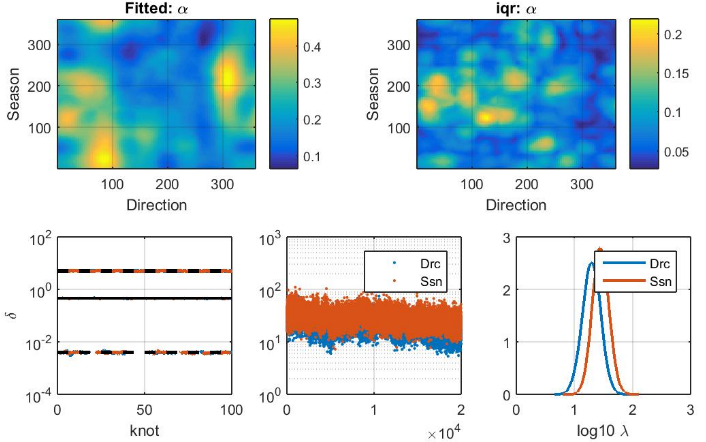
Parameters ψ



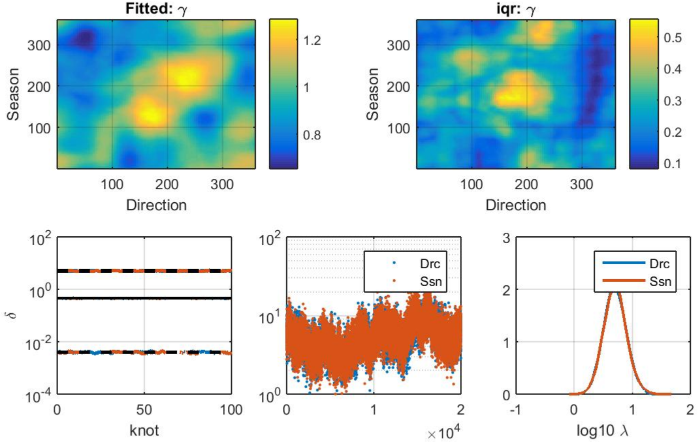
Parameters ρ



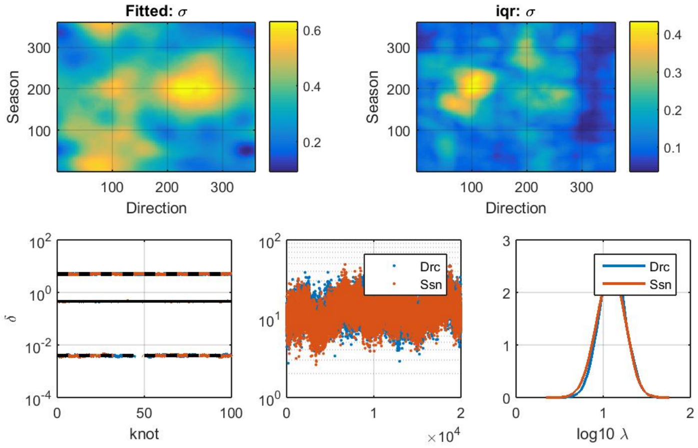
Parameters α



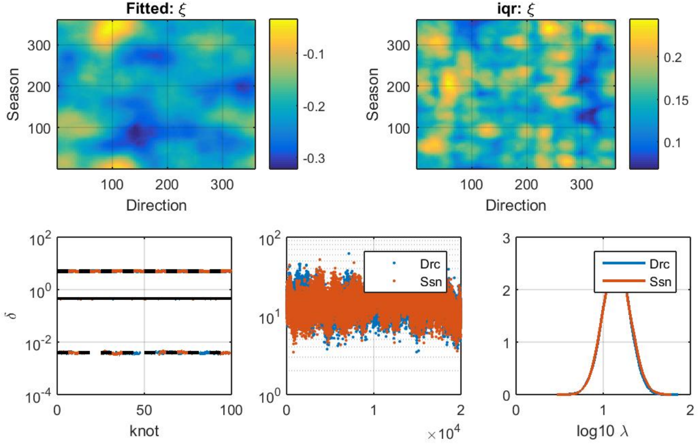
Parameters γ



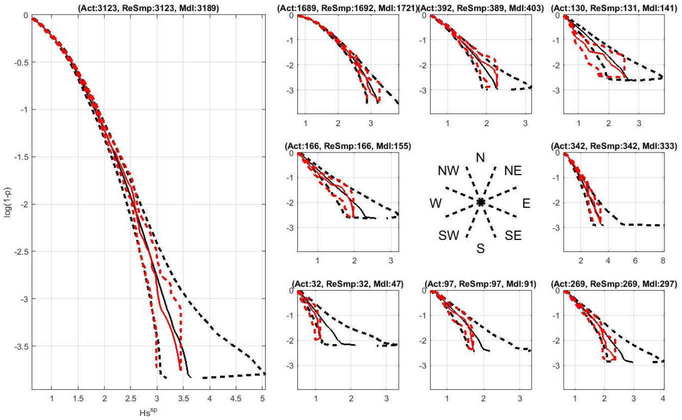
Parameters σ



Parameters ξ



Directional Seasonal 100Yr Return Values



Gradient Based MCMC

- **HMC**: Hamiltonian Monte Carlo: uses first derivatives of parameters have momentum based on gradient. This approach can be unstable so several leapfrog steps are taken instead of single step.
- **Riemann manifold HMC**: uses second derivatives of parameters. Here 2 leapfrog steps are needs so this is computationally challenging
- **MALA** Metropolis adjusted Langevin algorithm: uses first derivatives steps. Proposal calulated $\boldsymbol{\alpha}^*$ by sampling from a Normal distribution $N(\mu, \Sigma)$ where

$$\begin{aligned} \mu &= \boldsymbol{\alpha} - \frac{\epsilon}{2}\frac{\partial}{\partial \boldsymbol{\alpha}}(L + L_{prior}) \\ \Sigma &= \epsilon I \end{aligned} \tag{1}$$

and then implement standard MH based on this proposal.

mMALA

- Given a current state $\boldsymbol{\alpha}$ a proposal $\boldsymbol{\alpha}^*$ is sampled from $N(\mu(\boldsymbol{\alpha}), \Sigma)$, where

$$\begin{aligned} \mu(\boldsymbol{\alpha}) &= \boldsymbol{\alpha} - \frac{\epsilon}{2} G^{-1}(\boldsymbol{\alpha}) \frac{\partial}{\partial \boldsymbol{\alpha}} (L + L_{prior}) \\ \Sigma &= \epsilon G^{-1}(\boldsymbol{\alpha}) \end{aligned} \tag{2}$$

  and then MH is carried through as before. As in MALA we again do not have symmetric proposals and so we must calculate the full acceptance probability.

- it is also interesting to notice the similarities between IWLS and mMALA. To see this compare

$$G(\boldsymbol{\alpha}_\xi)^{-1} = (B' \frac{\partial^2 L}{\partial \boldsymbol{\xi}^2} B + \lambda_\xi P)^{-1} \tag{3}$$

$$\hat{\alpha}_{t+1} = (B' \hat{W}_t B + \lambda D' D)^{-1} B' \hat{W}_t \hat{z}_t \tag{4}$$