Biometrics in the Bush Capital

Using a linear mixed model based wavelet transform to model non-smooth trends arising from designed experiments

Clayton Forknall
Alison Kelly Yoni Nazarathy Ari Verbyla

26th November 2025

The University of Queensland
Australia

What's in a title?

Non-smooth trends arising from designed experiments

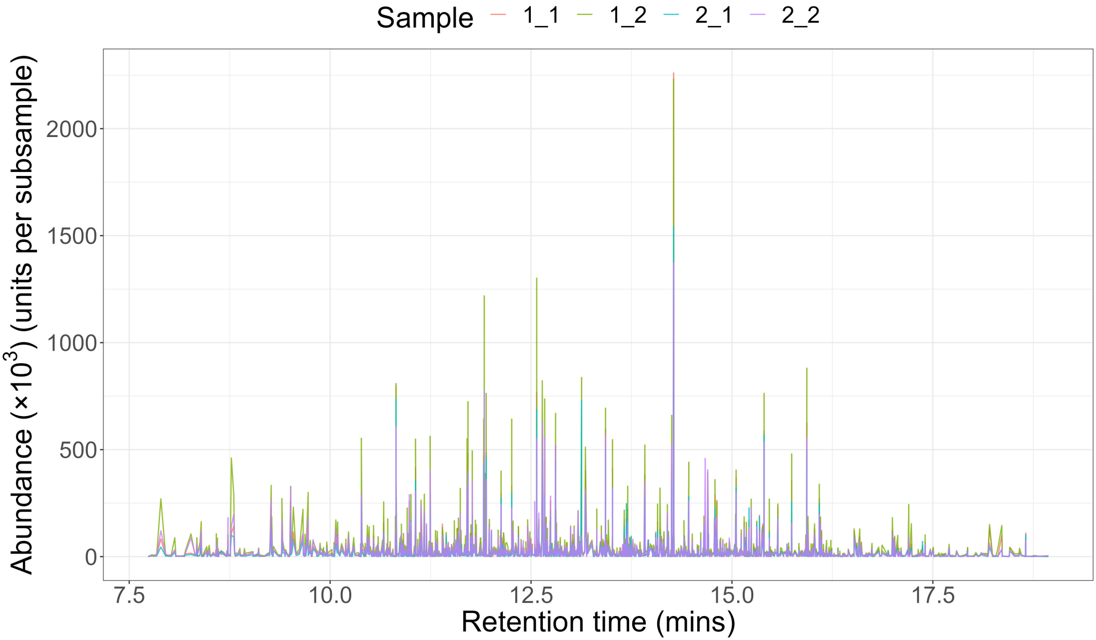
Linear mixed model based wavelet transform

Using LMM based wavelet transform to model non-smooth trends

A motivating example

Mass spectrometry (MS) based barley malt proteomics experiment (Forknall et al., 2023)

- **Multi-phase experiment (Brien & Bailey, 2006)**:
  - *Phase 1: Malt Sample Collection*
    - Two separate grain samples collected at commencement of malting processing ($g = 2$)
  - *Phase 2: MS Processing*
    - Two subsamples taken from each grain sample ($s = 2$)
    - Individual subsamples processed using MS based proteomics technique
- **Proteome composition via MS**
  - Same 1811 peptides detected/quantified from each subsample.
  - Peptides detected at unique, non-equidistant retention times ($t = 1811$).
  - Data set = 7,244 peptide abundance observations ($n = g\ s\ t = 7,244$).



Sample
1_1
1_2
2_1
2_2
Abundance ($\times 10^3$) (units per subsample)
0
500
1000
1500
2000
7.5
10.0
12.5
15.0
17.5
Retention time (mins)

What's in a title?

Non-smooth trends arising from designed experiments

Linear mixed model based wavelet transform

Using LMM based wavelet transform to model non-smooth trends

The linear mixed model (LMM) based wavelet transform

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{\mathrm{d}}\mathbf{u}_{\mathrm{d}} + \mathbf{Z}_{\mathrm{t}}\mathbf{u}_{w} + \mathbf{e}$$

- $\mathbf{y}$ is an $n \times 1$ vector of abundance observations.
- $\boldsymbol{\tau}$ is a vector of fixed effects, with associated design matrix $\mathbf{X}$.
- $\mathbf{u}_{\mathrm{d}}$ is a vector of random effects describing the experimental design structure, with associated design matrix $\mathbf{Z}_{\mathrm{d}}$.
- $\mathbf{e}$ is an $n \times 1$ vector of residual error effects.

The linear mixed model (LMM) based wavelet transform

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{\mathrm{d}}\mathbf{u}_{\mathrm{d}} + \mathbf{Z}_{\mathrm{t}}\mathbf{u}_{\omega} + \mathbf{e}$$

- $\mathbf{y}$ is an $n \times 1$ vector of abundance observations.
- $\boldsymbol{\tau}$ is a vector of fixed effects, with associated design matrix $\mathbf{X}$.
- $\mathbf{u}_{\mathrm{d}}$ is a vector of random effects describing the experimental design structure, with associated design matrix $\mathbf{Z}_{\mathrm{d}}$.
- $\mathbf{e}$ is an $n \times 1$ vector of residual error effects.

- $\mathbf{u}_{\omega}$ is a $t \times 1$ vector of random effects resulting from the LMM based wavelet transform.
- These effects describe the non-smooth response of abundance as a function of retention time.
- $\mathbf{Z}_{\mathrm{t}}$ is an $n \times t$ design matrix, necessary to respect multiple abundance observations at each retention time.

THE UNIVERSITY OF QUEENSLAND AUSTRALIA

The linear mixed model (LMM) based wavelet transform

$$\mathbf{y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{Z}_{\mathrm{d}}\mathbf{u}_{\mathrm{d}} + \mathbf{Z}_{\mathrm{t}}\mathbf{u}_{\omega} + \mathbf{e}$$

- Random and residual error effects are assumed to follow a normal distribution with a zero mean vector and variance-covariance matrix:

$$\mathrm{var}\left(\begin{bmatrix}\mathbf{u}_{\mathrm{d}}\\ \mathbf{u}_{\omega}\\ \mathbf{e}\end{bmatrix}\right) = \begin{bmatrix}\mathbf{G}_{\mathrm{d}} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_{\omega} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{R}\end{bmatrix}.$$

- Form of $\mathbf{u}_{\omega}$ and $\mathbf{G}_{\omega}$ is our focus and will be investigated further.

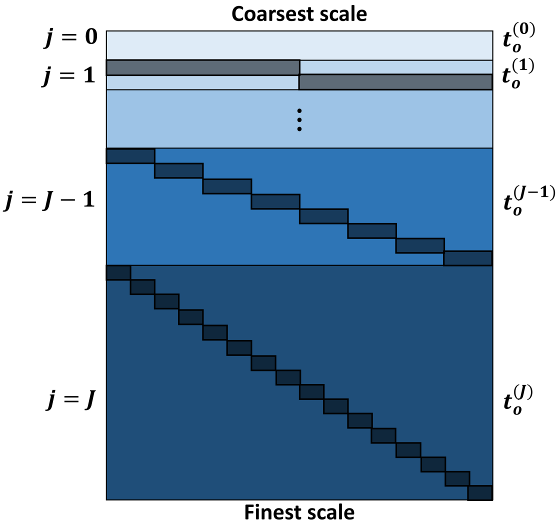
The wavelet transform

- *What is it?*
  - Mathematical construct proven to model non-smooth data, containing discontinuities.
- *How can I use it?*
  - Classical wavelets:
    - Built on framework of multiresolution analysis - relies on the Fourier transform.
    - Only applicable where observations are equidistant and dyadic ($\log_2(n)$ is integer) in number.
  - Second generation wavelets:
    - Implemented via 'lifting scheme' - retains multiscale properties of the classical transform.
    - Can be applied to non-equidistant observations, of any number.
- *Has it been incorporated into the LMM framework before?*
  - Classical wavelets:
    - Yes - Morris & Carroll, (2006); Wand & Ormerod, (2011).
  - Second generation wavelets:
    - Not until now!

The second generation wavelet transform

Wavelet scales

- Wavelet transform provides representation of data/effects at a series of **scales**.
- $J = \lceil \log_2(t) \rceil$ corresponds to the number of scales in the wavelet transform.
- $t_o^{(j)}$ is the number of wavelet functions/coefficients at each scale j.
- As j increases, $t_o^{(j)}$ increases, but support of wavelet functions decreases.
- This formulation allows for representation of non-smooth trends, as influence of spikes limited in terms of scale and location (Nason, 2008).
- Wavelet transform can be implemented through wavelet basis matrix, $\mathbf{W}^{-1}$.

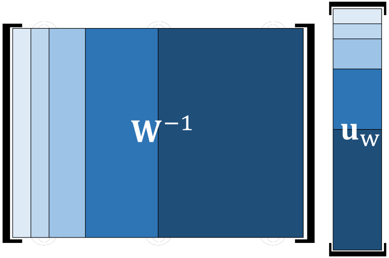


The second generation wavelet transform

Form of $\mathbf{u}_\omega$

$$\mathbf{u}_\omega = \mathbf{W}^{-1}\mathbf{u}_\mathrm{w}$$

- $\mathbf{W}^{-1} = \begin{bmatrix} \boldsymbol{\phi}^{(0)} & \boldsymbol{\Psi}^{(0)} & \boldsymbol{\Psi}^{(1)} & \dots & \boldsymbol{\Psi}^{(J-1)} \end{bmatrix}$ is the $t \times t$ wavelet basis matrix.
  - $\boldsymbol{\phi}^{(0)}$ and $\boldsymbol{\Psi}^{(j)}$ are $t \times t_\mathrm{o}^{(j)}$ matrices containing the values of the wavelet scaling function, $\phi^{(0)}(x)$, and wavelet functions, $\psi^{(j)}(x)$, at scale j.
- $\mathbf{u}_\mathrm{w} = \begin{bmatrix} \boldsymbol{\varphi}\ \boldsymbol{\omega}^{(0)\top} & \boldsymbol{\omega}^{(1)\top} & \dots & \boldsymbol{\omega}^{(J-1)\top} \end{bmatrix}^\top$ is the $t \times 1$ vector of wavelet coefficients.
  - $\boldsymbol{\varphi}$ and $\boldsymbol{\omega}^{(j)}$ are $t_\mathrm{o}^{(j)} \times 1$ vectors of coefficients associated with the wavelet scaling function and wavelet functions at scale j.

$\mathbf{u}_\omega =$ 

The second generation wavelet transform

Form of $\mathbf{G}_{\omega}$

$$\mathbf{u}_{\omega} = \mathbf{W}^{-1}\mathbf{u}_{\mathrm{w}}$$

Two options for $\mathrm{var}\left(\mathbf{u}_{\mathrm{w}}\right)$:

1. **Simple wavelet transform**:
   - Simple variance component controls extent of 'non-smoothness'.
   - $\mathbf{u}_{\mathrm{w}} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2_{\mathrm{w}}\mathbf{I}_t\right)$
   - $\mathbf{G}_{\omega} = \sigma^2_{\mathrm{w}}\mathbf{W}^{-1}{\mathbf{W}^{-1}}^{\top}$

$$\mathbf{u}_{\omega} = \mathbf{W}^{-1}\mathbf{u}_{\mathrm{w}}$$

Two options for $\operatorname{var}(\mathbf{u}_{\mathrm{w}})$:

2 **Partitioned wavelet transform**:

- Uses wavelet scale structure implicit in wavelet functions, allowing for heterogeneous wavelet variance at each wavelet scale:

$$\operatorname{var}(\mathbf{u}_{\mathrm{w}}) = \operatorname{var}\left(\begin{bmatrix} \varphi \\ \boldsymbol{\omega}^{(0)} \\ \boldsymbol{\omega}^{(1)} \\ \vdots \\ \boldsymbol{\omega}^{(J-1)} \end{bmatrix}\right) = \begin{bmatrix} \sigma_{\varphi}^2 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_{\omega(0)}^2 \mathbf{I}_{t_{\mathrm{o}}^{(0)}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_{\omega(1)}^2 \mathbf{I}_{t_{\mathrm{o}}^{(1)}} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \sigma_{\omega(J-1)}^2 \mathbf{I}_{t_{\mathrm{o}}^{(J-1)}} \end{bmatrix}$$

- $\mathbf{G}_{\omega} = \sigma_{\varphi}^2 \boldsymbol{\phi}^{(0)} \boldsymbol{\phi}^{(0)\top} + \sum_{j=0}^{J-1} \sigma_{\omega(j)}^2 \boldsymbol{\Psi}^{(j)} \boldsymbol{\Psi}^{(j)\top}$

What's in a title?

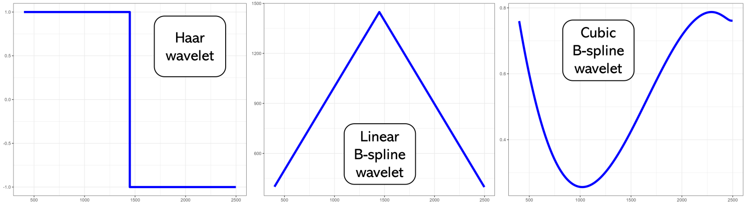
Non-smooth trends arising from designed experiments

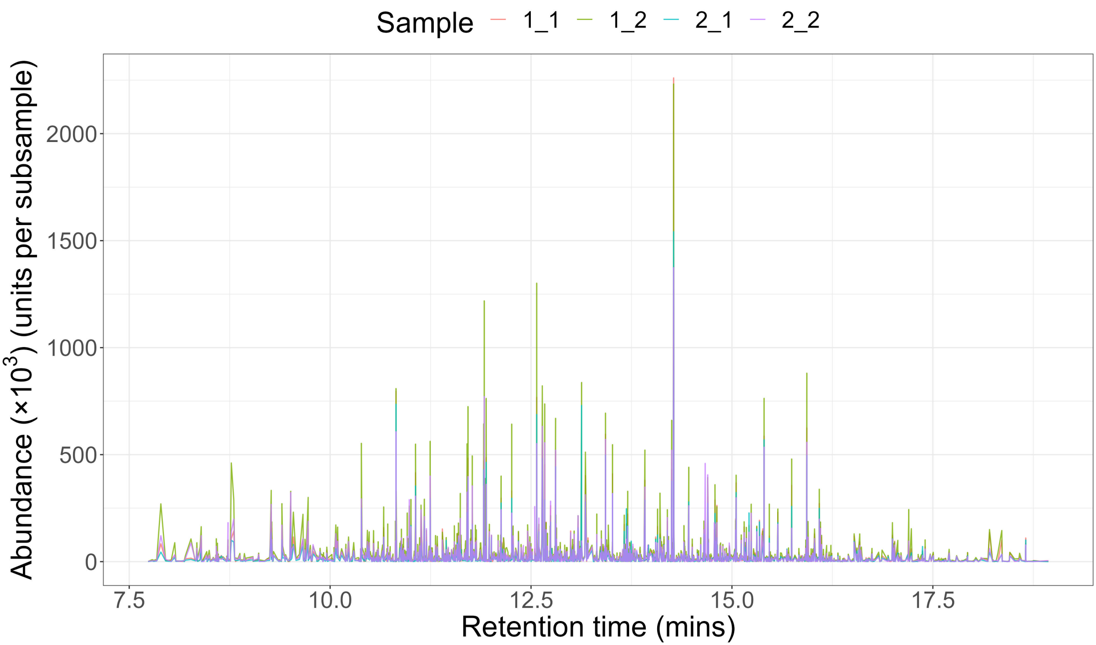
Linear mixed model based wavelet transform

Using LMM based wavelet transform to model non-smooth trends

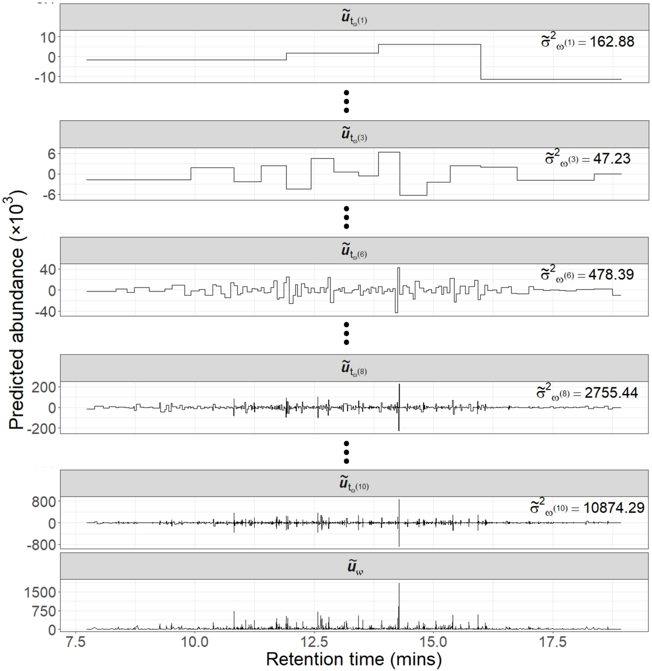
Implementing LMM based wavelet transform
Software solutions

- **LMM implementation**:
  - Implementable in any LMM software or package that enables the user to specify their own design matrix, or matrix of basis functions.
  - Successfully implemented in `R` using:
    - `asreml` (The VSNi Team, 2023),
    - `lmmSolver` (Boer, 2023),
    - `sommer` (Covarrubias-Pazaran, 2016).
- **Second generation wavelet basis construction**
  - I developed an `R` package to construct the B-spline wavelet basis matrix (Jansen 2016, 2022).
  - Package calculates $\mathbf{W}^{-1}$ for:
    - Haar wavelet
    - Linear B-spline wavelet
    - Cubic B-spline wavelet





Sample
1_1
1_2
2_1
2_2
Abundance (×10^3) (units per subsample)
2000
1500
1000
500
0
7.5
10.0
12.5
15.0
17.5
Retention time (mins)



Wavelet variances

j	$\tilde{\sigma}^2_{\omega(j)}$
0	0.00
1	**162.88**
2	71.11
3	**47.23**
4	248.52
5	486.20
6	**478.39**
7	1454.10
8	**2755.44**
9	5651.23
10	**10874.29**

Future work

- Explore contribution of different scales of the wavelet transform - are all scales necessary?
- Potential to explore two dimensional setting - tensor wavelet transforms.
- Forthcoming methodology paper.

Acknowledgements

- Supervisors:
  - Associate Professor Yoni Nazarathy
  - Dr Alison Kelly
  - Professor Ari Verbyla
- Institutions:
  - The University of Queensland
  - Queensland Department of Primary Industries
  - The University of Adelaide, in particular The Plant Accelerator, part of the APPN
- Funding:
  - The University of Queensland via a Research Training Program Tuition Fee Offset
  - Queensland Department of Primary Industries

References

Boer, M. P. 2023. Tensor product P-splines using a sparse mixed model formulation. *Statistical Modelling*, 23: 465-479.

Brien, C. J. and Bailey, R. A. 2006. Multiple randomisations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68: 571 - 609.

Covarrubias-Pazaran, G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE*, 11: 1-15.

Forknall, C. R., Verbyla, A. P., Nazarathy, Y., Yousif, A., Osama, S., Jones, S. H., Kerr, E., Schulz, B. L., Fox, G. P. and Kelly, A. M. 2023. Covariance Clustering: Modelling covariance in designed experiments when the number of variables is greater than experimental units. *Journal of Agricultural, Biological and Environmental Statistics*, 29 (2): 232-256.

Jansen, M. 2016. Non-equispaced B-spline wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, 14 (6).

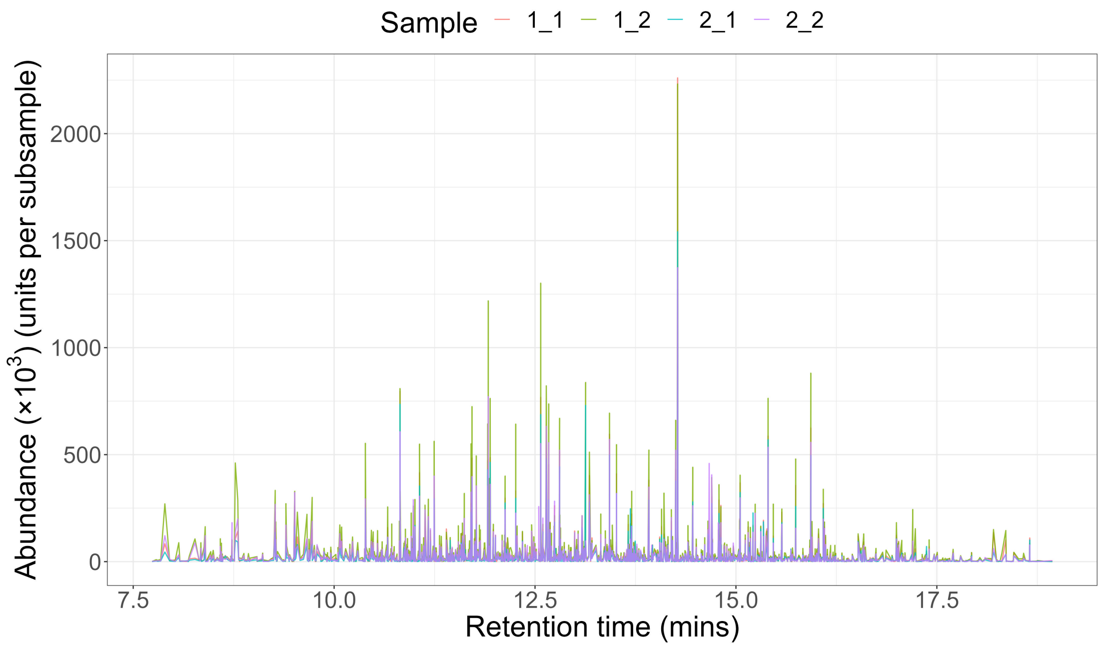
Jansen, M. 2022. *Wavelets from a Statistical Perspective*. Book. Boca Raton: CRC Press.

Morris, J. S. and Carroll, R. J. 2006. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68: 179 - 199.

Nason, G. P. 2008. *Wavelet Methods in Statistics with R*. Book. New York: Springer.

The VSNi Team. 2023. asreml: Fits linear mixed models using REML. 4.2.0.370.

Wand, M. P. and Ormerod, J. T. 2011. Penalized wavelets: Embedding wavelets into semiparametric regression. *Electronic Journal of Statistics*, 5: 1654 - 1717.



Sample
1_1
1_2
2_1
2_2
Abundance (×10^3) (units per subsample)
0
500
1000
1500
2000
7.5
10.0
12.5
15.0
17.5
Retention time (mins)