A more smoothed fitted curve method: Cubic Spline with beta-knots for Thermal Shift Analysis in R (TSAR)

In the following text, "beta" or "beta method" is referred to Cubic Spline with beta-knots for Thermal Shift Analysis in R (TSAR).

Summary

This PR introduces a new smoothing method, beta method, overwriting for the `TSA_average()` function that utilizes a beta-knots natural cubic spline centered at numerical T_m point to fit a smoother curve to the melting point data. The new method exhibits **solves the problem of ossilating fitted curves at the boundaries resulted by General Additive Model (GAM).**

The new method is a overwritten method to exisiting method `TSA_average()` with two children methods calling the new method: `TSA_compare_plot()` and `TSA_wells_plot()` such that

- **`TSA_average()`**
  - New argument `smoother = c("gam", "beta", "none")` .
  - **Beta-knot natural cubic spline** option with interior knots drawn from **Beta(a, a)** and **centered at the condition's** T_m.
- **`TSA_compare_plot()`**
  - Accepts and forwards `smoother` to `TSA_average()` .
- **`TSA_wells_plot()`**
  - Accepts and forwards `smoother` to `TSA_average()` .

Motivation

Limitations of GAM

1. **Oscillation at Boundaries**: GAM often induces Runge's phenomenon, where fitted curves show oscillation at the boundaries, reducing its capacity to fit the data accurately and leading to misleading its interpretations.
2. **Lack of Domain Awareness**: GAM does not concentrate flexibility near the central melting-point region, resulting a different T_m from numerical T_m.

Resolution

The **Beta Method** uses a natural cubic spline with interior knots drawn from a Beta distribution centered interest's numerical T_m calculated priorly. This method stabilizes the fitted curve in boundary condition and in the same time provides a more accurate fit to the melting point data.

The idea of using Beta Method is drawn from the facts that (1) cubic spline is a well-known numerical method for smoothing data requiring few computations as well as providing a good numerical accuracy; (2) a beta distribution is a centered hill-like distribution that fulfills the demand of concentrated knot placements near the melting point region and sparse knots at the boundaries.

What’s Changed

1) `TSA_average()`

```r
TSA_average(
  tsa_data,
  y = c("Fluorescence", "RFU"),
  digits = 1,
  avg_smooth = TRUE,
  sd_smooth  = TRUE,
  smoother   = c("gam", "beta", "none"),
  beta_shape       = 3,       # a in Beta(a, a)
  beta_n_knots     = NULL,    # if NULL, use beta_knots_frac
  beta_knots_frac = 0.008,    # fraction of unique temperatures used as interior knots
  use_natural      = TRUE     # natural cubic (recommended)
)
```

> Note: `smoother="gam"` reproduces previous output.

2) `TSA_compare_plot()`

- New argument: `smoother = c("gam", "beta", "none")`

3) `TSA_wells_plot()`

- New argument: `smoother = c("gam", "beta", "none")`

What’s New

`TSA_smoother_diagnostics()`

- A diagnostic method to tests the fitted curves with different smoothing methods and calculates errors. It intends to help users

to find the best smoothing method/parameters with lowest error.
- It will return a data frame with the following columns:
  - `condition_id` : the condition ID
  - `method` : candidate smoothing method used
  - `error` : the error of the fitted curve to the original data
- It will also print a report of the best method/parameters with the lowest error.

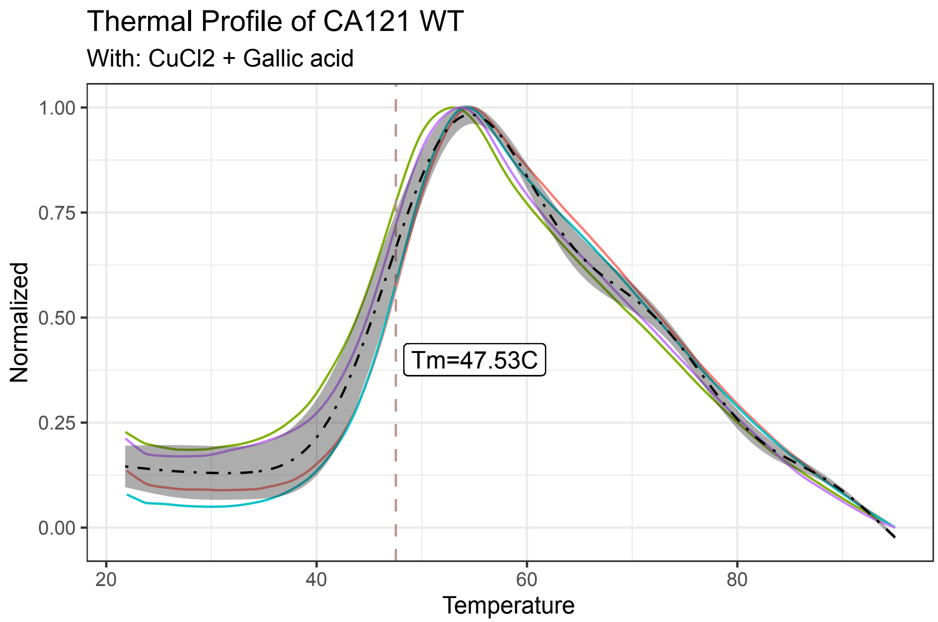
```
TSA_smoother_diagnostics <- function(
    tsa_data,
    y = c("Fluorescence", "RFU"),
    metric = c("L2", "L1", "R2"),
    # beta grids to test; if NULL or length 0, default to a=4 and frac=0.008
    beta_shapes = NULL,
    beta_fracs  = NULL,
    # pass-through options if you need them
    digits = 1,
    use_natural = TRUE,
    beta_n_knots = NULL
)
```

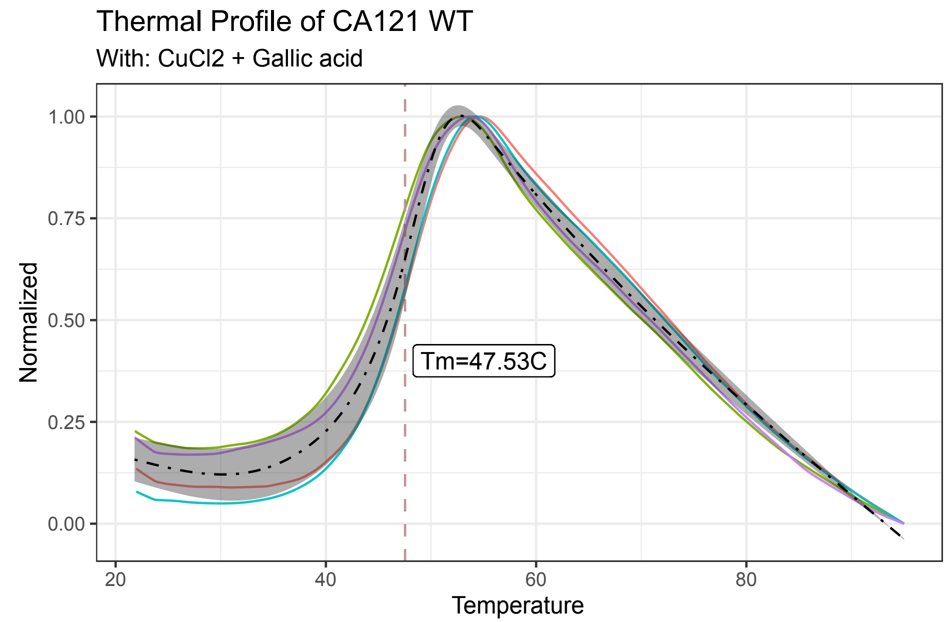
Performance

> Dotted curves represent the fitted curve; colored curves represent the data.

Example 1 Beta method shows smoother fitted curves than GAM

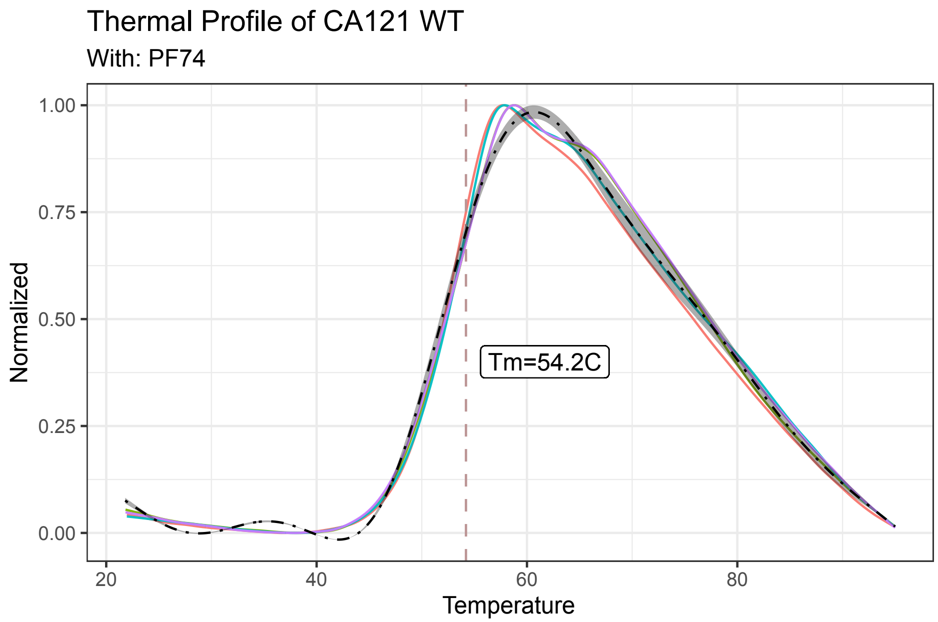
GAM	Beta Method

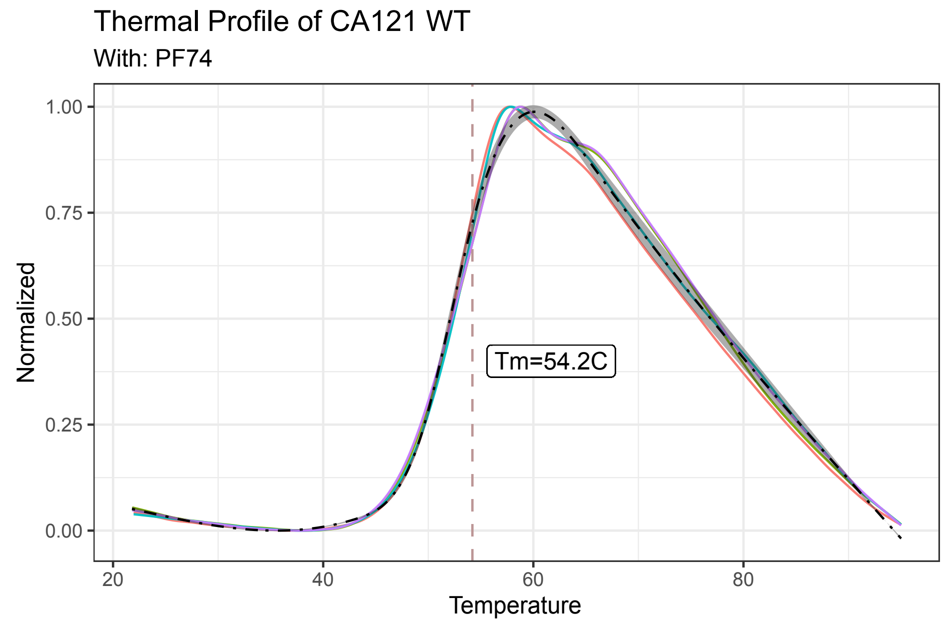




Example 2 Beta method shows less ossilating boundary condition than GAM

GAM	Beta Method





Example 3 With 13 samples in one experiment of TSA, Beta method shows more accurate fitted curves (~30.4% improvement) than GAM

condition_id	method	error
CA121 WT_dH2O	gam	0.03883613
CA121 WT_dH2O	beta(a=4, frac=0.008)	0.01457599
CA121 WT_PF74	gam	0.02145768
CA121 WT_PF74	beta(a=4, frac=0.008)	0.01469731
CA121 WT_Gallic acid	gam	0.02233872
CA121 WT_Gallic acid	beta(a=4, frac=0.008)	0.02583934
CA121 WT_FeCl3 + Gallic acid	gam	0.02017039

CA121 WT_FeCl3 + Gallic acid	beta(a=4, frac=0.008)	0.01009895
CA121 WT_CuCl2 + Gallic acid	gam	0.01720631
CA121 WT_CuCl2 + Gallic acid	beta(a=4, frac=0.008)	0.01629382
CA121 WT_CaCl2 + Gallic acid	gam	0.01285055
CA121 WT_CaCl2 + Gallic acid	beta(a=4, frac=0.008)	0.03252719
CA121 WT_MgCl2 + Gallic acid	gam	0.02231318
CA121 WT_MgCl2 + Gallic acid	beta(a=4, frac=0.008)	0.02593790
CA121 WT_NaCl + Gallic acid	gam	0.02344820
CA121 WT_NaCl + Gallic acid	beta(a=4, frac=0.008)	0.02589357
CA121 WT_FeCl3	gam	0.03972613
CA121 WT_FeCl3	beta(a=4, frac=0.008)	0.01665110
CA121 WT_CuCl2	gam	0.03298392
CA121 WT_CuCl2	beta(a=4, frac=0.008)	0.01408443
CA121 WT_CaCl2	gam	0.04074613
CA121 WT_CaCl2	beta(a=4, frac=0.008)	0.02054448
CA121 WT_MgCl2	gam	0.03869195
CA121 WT_MgCl2	beta(a=4, frac=0.008)	0.01902785
CA121 WT_NaCl	gam	0.03990452

CA121 WT_NaCl	beta(a=4, frac=0.008)	0.02183548

mean L2 error of gam method: 0.02851
mean L2 error of beta(a=4, frac=0.008): 0.01985

beta(a=4, frac=0.008) has lowest error and highest accuracy.

> The table and report above was generated by `TSA_smoother_diagnostics()`

Usage

see `demo.Rmd`