smoothSDE vignette

Théo Michelot

2020-11-26

The R package smoothSDE implements varying-coefficient versions of several popular stochastic differential equations (SDEs). It can be used to estimate linear or non-linear relationships between SDE parameters and time-varying covariates. It also provides functions for model visualisation, uncertainty estimation, and residual analysis.

1 Package description

1.1 Varying-coefficient stochastic differential equations

Varying-coefficient SDEs, and the inferential method implemented in smoothSDE, are described by Michelot et al. (2020). The main model formulations currently implemented in the package are described in the following table.

Model	Name	Parameters		Link functions
Brownian motion	"BM"	drift	$\mu \in (-\infty, +\infty)$	identity
		diffusion	$\sigma > 0$	log
Ornstein-Uhlenbeck process	"OU"	mean	$\mu \in (-\infty, +\infty)$	identity
		reversion	$\beta > 0$	log
		diffusion	$\sigma > 0$	log
Continuous-time correlated random walk	"CTCRW"	reversion	$\beta > 0$	log
		diffusion	$\sigma > 0$	log

Brownian motion and the Ornstein-Uhlenbeck process have a wide range of applications (e.g., ecology, finance), and the continuous-time correlated random walk is a popular model of animal movement described by Johnson et al. (2008).

1.2 SDE class

The package is centred around the R6 class *SDE*, which encapsulates the model formulas and the data for a varying-coefficient SDE model. The constructor SDE$new can be used to create a model object, and its main arguments are:

- formulas: List of formulas for the parameters of the SDE. The formulas can include terms from standard R expressions, as well as smooth terms and random effects from the mgcv package (Wood (2017)).
- data: Data frame with columns for the response variable, for the covariate, for ID, and for time.
- type: Type of SDE; options are "BM" (Brownian motion), "OU" (Ornstein-Uhlenbeck), "CTCRW" (continuous-time correlated random walk).
- response: Name of response variable(s).

2 Example: elephant movement analysis

The following code can be used to fit the varying-coefficient CTCRW model to elephant GPS data from Wall et al. (2014), as described in Section 3.2 of Michelot et al. (2020).

We first download the data from the Movebank data repository, and keep the relevant rows and columns. Note that the CTCRW model requires projected Easting-Northing locations (rather than longitude-latitude), so that's what we are working with here.

```
# Load data and keep relevant columns
URL <- paste0("https://www.datarepository.movebank.org/bitstream/handle/",
              "10255/move.373/Elliptical%20Time-Density%20Model%20%28Wall%",
              "20et%20al.%202014%29%20African%20Elephant%20Dataset%20%",
              "28Source-Save%20the%20Elephants%29.csv")
raw <- read.csv(url(URL))
keep_cols <- c(11, 13, 14, 17, 4, 5, 6)
raw_cols <- raw[, keep_cols]
colnames(raw_cols) <- c("ID", "x", "y", "date", "lon", "lat", "temp")

# Only keep five months to eliminate seasonal effects
track <- subset(raw_cols, ID == unique(ID)[1])
dates <- as.POSIXlt(track$date, tz = "GMT")
times <- as.numeric(dates - min(dates))/3600
keep_rows <- which(dates > as.POSIXct("2009-05-01 00:00:00") &
                       dates < as.POSIXct("2009-09-30 23:59:59"))
track <- track[keep_rows,]
dates <- dates[keep_rows]
times <- times[keep_rows]

# Convert to km
track$x <- track$x/1000
track$y <- track$y/1000
```

We create a data frame that includes columns for

- time series identifier "ID". This is required when fitting the model to several time series, treated as independent realisations of the same underlying process. Here, we only have one time series, so all rows have the same ID.
- responses "x" and "y". Here, there are two response variables because the CTCRW model works in two dimensions (Easting and Northing). For other models, such as Brownian motion and Ornstein-Uhlenbeck process, only one response variable is expected.
- covariate "temp". We will later estimate the effect of this covariate on the parameters of the elephant's velocity process.
- time. This should be a numeric column for time, used to compute time intervals between observations.

```
data <- data.frame(ID = 1,
                   x = track$x,
                   y = track$y,
                   temp = track$temp,
                   time = times)

head(data)
```

```
  ID        x        y temp time
1  1 572.3427 1675.424   33 9703
```

```
2   1 572.5443 1675.392   32 9704
3   1 572.6159 1675.339   31 9705
4   1 572.7745 1675.101   31 9706
5   1 572.8844 1675.065   31 9707
6   1 573.6659 1674.322   30 9708
```

In this analysis, we want to look into the effects of external temperature on the elephant's movement. We specify this by expressing both parameters of the CTCRW model (velocity reversion β and velocity diffusion σ) as smooth functions of the temperature. For these smooth terms, we use the syntax from the R package mgcv; here defining cubic regression splines with 10 basis functions and with shrinkage. See the mgcv documentation for more details.

```
formulas <- list(beta = ~ s(temp, k = 10, bs = "cs"),
                 sigma = ~ s(temp, k = 10, bs = "cs"))
```

Finally, we use the constructor of the SDE class to create an object for this model, passing as input the formulas, data, type of model, and name of response variables.

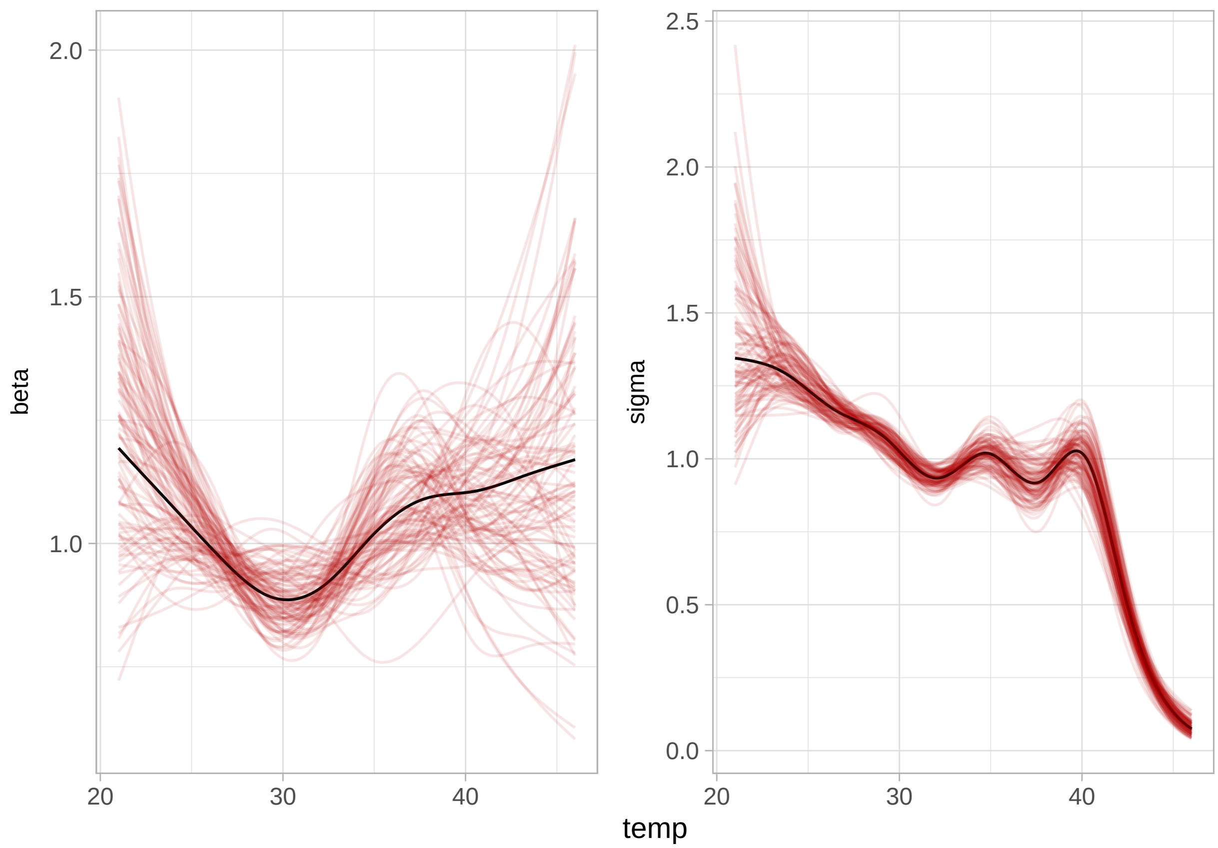
```
my_sde <- SDE$new(formulas = formulas,
                  data = data,
                  type = "CTCRW",
                  response = c("x", "y"))
```

Once the model has been created, the parameters can be estimated using the method "fit":

```
my_sde$fit()
```

Finally, the relationship between the SDE parameters and the covariates can be plotted with the method "plot_par". The function can also generate posterior samples for the estimated relationship, to visualise uncertainty. Here, we plot the CTCRW parameters as function of temperature, with 100 posterior samples:

```
my_sde$plot_par("temp", n_post = 100)
```



As described by Michelot et al. (2020), these results suggest that this elephant tended to move less at high temperatures, with a big drop in speed over 40 degrees Celsius.

Estimates and point-wise confidence intervals of the SDE parameters can also be computed for given covariate values, using the predict_par method.

```r
# Data frame with covariate values for prediction
new_data <- data.frame(temp = c(20, 30, 40))

# Get estimates and CIs
par <- my_sde$predict_par(new_data = new_data, CI = TRUE)

# Estimates and 95% CIs for each temperature value given as input
par
```

```
$estimate
         beta    sigma
[1,] 1.234930 1.354184
[2,] 0.886193 1.024858
[3,] 1.103145 1.019070

$low
          beta     sigma
[1,] 0.7943178 0.8360959
[2,] 0.7766455 0.9561926
[3,] 0.9256706 0.8936987
```

```
$upp
         beta    sigma
[1,] 1.918143 2.267873
[2,] 1.019487 1.093547
[3,] 1.313710 1.160171
```

3 To-do list

```
- Non-zero mean velocity in CTCRW model

- Two-dimensional OU process

- One-dimensional CTCRW process

- Implement decaying-response model
```

References

Johnson, Devin S, Joshua M London, Mary-Anne Lea, and John W Durban. 2008. "Continuous-Time Correlated Random Walk Model for Animal Telemetry Data." *Ecology* 89 (5): 1208–15.

Michelot, Théo, Richard Glennie, Catriona Harris, and Len Thomas. 2020. "Varying-Coefficient Stochastic Differential Equations with Applications in Ecology." *arXiv Preprint arXiv:2008.09111*.

Wall, Jake, George Wittemyer, Valerie LeMay, Iain Douglas-Hamilton, and Brian Klinkenberg. 2014. "Elliptical Time-Density Model to Estimate Wildlife Utilization Distributions." *Methods in Ecology and Evolution* 5 (8): 780–90.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. CRC press.