

# Package ‘Rdrw’

May 8, 2026

**Version** 1.0.2

**Date** 2020-9-8

**Title** Univariate and Multivariate Damped Random Walk Processes

**Author** Zhirui Hu and Hyungsuk Tak

**Maintainer** Hyungsuk Tak <hyungsuk.tak@gmail.com>

**Depends** R (>= 2.2.0)

**Imports** mvtnorm(>= 1.0-11)

**Description** We provide a toolbox to fit and simulate a univariate or multivariate damped random walk process that is also known as an Ornstein-Uhlenbeck process or a continuous-time autoregressive model of the first order, i.e., CAR(1) or CARMA(1, 0). This process is suitable for analyzing univariate or multivariate time series data with irregularly-spaced observation times and heteroscedastic measurement errors. When it comes to the multivariate case, the number of data points (measurements/observations) available at each observation time does not need to be the same, and the length of each time series can vary. The number of time series data sets that can be modeled simultaneously is limited to ten in this version of the package. We use Kalman-filtering to evaluate the resulting likelihood function, which leads to a scalable and efficient computation in finding maximum likelihood estimates of the model parameters or in drawing their posterior samples. Please pay attention to loading the data if this package is used for astronomical data analyses; see the details in the manual. Also see Hu and Tak (2020) <[doi:10.48550/arXiv.2005.08049](https://doi.org/10.48550/arXiv.2005.08049)>.

**License** GPL-2

**Encoding** UTF-8

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2020-09-08 17:50:02 UTC

## Contents

drw	2
drw.sim	7
Rdrw	9

<b>Index</b>	<b>11</b>
--------------	-----------

**Description**

The function `drw` fits univariate and multivariate damped random walk processes on multiple time series data sets possibly with known measurement error standard deviations via state-space representation. This function `drw` evaluates the resulting likelihood function of the model parameters via Kalman-filtering whose minimum complexity is linear in the number of unique observation times. The function returns the maximum likelihood estimates or posterior samples of the model parameters. For astronomical data analyses, users need to pay attention to loading the data because R's default is to load only seven effective digits; see details below.

**Usage**

```
drw(data1, data2, data3, data4, data5,
    data6, data7, data8, data9, data10,
    n.datasets, method = "mle",
    bayes.n.burn, bayes.n.sample,
    mu.UNIFprior.range = c(-30, 30),
    tau.IGprior.shape = 1, tau.IGprior.scale = 1,
    sigma2.IGprior.shape = 1, sigma2.IGprior.scale = 1e-7)
```

**Arguments**

- |       |  |
|-------|--|
| data1 | An ( $n_1$ by 3) matrix for time series data 1. The first column has $n_1$ observation times, the second column contains $n_1$ measurements (magnitudes), the third column includes $n_1$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_1$ zeros.   |
| data2 | Optional if more than one time series data set are available. An ( $n_2$ by 3) matrix for time series data 2. The first column has $n_2$ observation times, the second column contains $n_2$ measurements/observations/magnitudes, the third column includes $n_2$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_2$ zeros.    |
| data3 | Optional if more than two time series data sets are available. An ( $n_3$ by 3) matrix for time series data 3. The first column has $n_3$ observation times, the second column contains $n_3$ measurements/observations/magnitudes, the third column includes $n_3$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_3$ zeros.   |
| data4 | Optional if more than three time series data sets are available. An ( $n_4$ by 3) matrix for time series data 4. The first column has $n_4$ observation times, the second column contains $n_4$ measurements/observations/magnitudes, the third column includes $n_4$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_4$ zeros. |

data5	Optional if more than four time series data sets are available. An ( $n_5$ by 3) matrix for time series data 5. The first column has $n_5$ observation times, the second column contains $n_5$ measurements/observations/magnitudes, the third column includes $n_5$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_5$ zeros.
data6	Optional if more than five time series data sets are available. An ( $n_6$ by 3) matrix for time series data 6. The first column has $n_6$ observation times, the second column contains $n_6$ measurements/observations/magnitudes, the third column includes $n_6$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_6$ zeros.
data7	Optional if more than six time series data sets are available. An ( $n_7$ by 3) matrix for time series data 7. The first column has $n_7$ observation times, the second column contains $n_7$ measurements/observations/magnitudes, the third column includes $n_7$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_7$ zeros.
data8	Optional if more than seven time series data sets are available. An ( $n_8$ by 3) matrix for time series data 8. The first column has $n_8$ observation times, the second column contains $n_8$ measurements/observations/magnitudes, the third column includes $n_8$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_8$ zeros.
data9	Optional if more than eight time series data sets are available. An ( $n_9$ by 3) matrix for time series data 9. The first column has $n_9$ observation times, the second column contains $n_9$ measurements/observations/magnitudes, the third column includes $n_9$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_9$ zeros.
data10	Optional if more than nine time series data sets are available. An ( $n_{10}$ by 3) matrix for time series data 10. The first column has $n_{10}$ observation times, the second column contains $n_{10}$ measurements/observations/magnitudes, the third column includes $n_{10}$ measurement error standard deviations. If the measurement error standard deviations are unknown, the third column must be a vector of $n_{10}$ zeros. The current version of package allows up to ten time series data sets. If users have more than ten time series data sets to be modeled simultaneously, please contact the authors.
n.datasets	The number of time series data sets to be modeled simultaneously. An integer value inclusively between 1 to 10. For example, if "n.datasets = 3", then users must enter data1, data2, and data3 as inputs of drw.
method	If method = "mle", it returns maximum likelihood estimates of the model parameters. If method = "bayes" it produces posterior samples of the model parameters.
bayes.n.burn	Required for method = "bayes". The number of warming-up iterations for a Markov chain Monte Carlo method.
bayes.n.sample	Required for method = "bayes". The size of a posterior sample for each parameter for a Markov chain Monte Carlo method.

`mu.UNIFprior.range`

Required for method = "bayes". The range of the Uniform prior on each long-term average  $\mu_j$  of the process, where  $j$  goes from 1 to the total number of time series data sets. The default range is  $(-30, 30)$  for astronomical applications.

`tau.IGprior.shape`

Required for method = "bayes". The shape parameter of the invserse-Gamma prior on each timescale  $\tau_j$  of the process, where  $j$  goes from 1 to the total number of time series data sets. The default shape parameter is one for astronomical applications.

`tau.IGprior.scale`

Required for method = "bayes". The scale parameter of the invserse-Gamma prior on each timescale  $\tau_j$ , where  $j$  goes from 1 to the total number of time series data sets. The default scale parameter is one for astronomical applications.

`sigma2.IGprior.shape`

Required for method = "bayes". The shape parameter of the invserse-Gamma prior on each short-term variability (variance)  $\sigma_j^2$ , where  $j$  goes from 1 to the total number of time series data sets. The default shape parameter is one for astronomical applications.

`sigma2.IGprior.scale`

Required for method = "bayes". The scale parameter of the invserse-Gamma prior on each short-term variability (variance)  $\sigma_j^2$ , where  $j$  goes from 1 to the total number of time series data sets. The default shape parameter is  $1e-7$  for astronomical applications.

## Details

The multivariate damped random walk process  $\mathbf{X}(t)$  is defined by the following stochastic differential equation:

$$d\mathbf{X}(t) = -D_\tau^{-1}(\mathbf{X}(t) - \boldsymbol{\mu})dt + D_\sigma d\mathbf{B}(t),$$

where  $\mathbf{X}(t) = \{X_1(t), \dots, X_k(t)\}$  is a vector of  $k$  measurements/observations/magnitudes of the  $k$  time series data sets in continuous time  $t \in R$ ,  $D_\tau$  is a  $k \times k$  diagonal matrix whose diagonal elements are  $k$  timescales with each  $\tau_j$  representing the timescale of the  $j$ -th time series data,  $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_k\}$  is a vector for long-term averages of the  $k$  time series data sets,  $D_\sigma$  is  $k \times k$  diagonal matrix whose diagonal elements are short-term variabilities (standard deviation) of  $k$  time series data sets, and finally  $\mathbf{B}(t) = \{B_1(t), \dots, B_k(t)\}$  is a vector for  $k$  standard Brownian motions whose  $k(k-1)/2$  pairwise correlations are modeled by correlation parameters  $\rho_{jl}$  ( $1 \leq j < l \leq k$ ) such that  $dB_j(t)B_l(t) = \rho_{jl}dt$ .

We evaluate this continuous-time process at  $n$  discrete observation times  $\mathbf{t} = \{t_1, \dots, t_n\}$ . The observed data  $\mathbf{x} = \{x_1, \dots, x_n\}$  are multiple time series data measured at irregularly spaced observation times  $\mathbf{t}$  with possibly known measurement error standard deviations,  $\boldsymbol{\delta} = \{\delta_1, \dots, \delta_n\}$ . Since one or more time series observations can be measured at each observation time  $t_i$ , the length of a vector  $x_i$  can be different, depending on how many time series observations are available at the  $i$ -th observation time. We assume that these observed data  $\mathbf{x}$  are realizations of the latent time series data sets  $\mathbf{X}(\mathbf{t}) = \{\mathbf{X}(t_1), \dots, \mathbf{X}(t_n)\}$  with Gaussian measurement errors whose variances are  $\boldsymbol{\delta}$ . This is a typical setting of state-space modeling. We note that if the measurement error variances are unknown,  $\boldsymbol{\delta}$  must be set to zeros, which means that the observed data directly measure the latent values.

Please note that when astronomical time series data are loaded on R by `read.table`, `read.csv`, etc., some decimal places of the the observation times are automatically rounded because R's default is to load seven effective digits. For example, R will load the observation time 51075.412789 as 51075.41. This default will produce many ties in observation times even though there is actually no tie in observation times. To prevent this, please type `options(digits = 11)` before loading the data if the observation times are in seven effective digits.

## Value

The outcomes of `drw` are composed of:

- mu** The maximum likelihood estimate(s) of the long-term average(s) if method is "mle", and the posterior sample(s) of the long-term average(s) if method is "bayes". In the former case (mle), it is a vector of length  $k$ , where  $k$  is the number of time series data sets used. In the later case (bayes), it is an  $(m$  by  $k)$  matrix where  $m$  is the size of the posterior sample.
- sigma** The maximum likelihood estimate(s) of the short-term variability (standard deviation) parameter(s) if method is "mle", and the posterior sample(s) of the short-term variability parameter(s) if method is "bayes". In the former case (mle), it is a vector of length  $k$ , where  $k$  is the number of time series data sets used. In the later case (bayes), it is an  $(m$  by  $k)$  matrix where  $m$  is the size of the posterior sample.
- tau** The maximum likelihood estimate(s) of the timescale(s) if method is "mle", and the posterior sample(s) of the timescale(s) if method is "bayes". In the former case (mle), it is a vector of length  $k$ , where  $k$  is the number of time series data sets used. In the later case (bayes), it is an  $(m$  by  $k)$  matrix where  $m$  is the size of the posterior sample.
- rho** Only when more than one time series data set are used. The maximum likelihood estimate(s) of the (cross-) correlation(s) if method is "mle", and the posterior sample(s) of the (cross-) correlation(s) if method is "bayes". In the former case (mle), it is a vector of length  $k(k-1)/2$ , where  $k$  is the number of time series data sets used, i.e.,  $\rho_{12}, \dots, \rho_{1k}, \rho_{23}, \dots, \rho_{2k}, \dots, \rho_{k-1,k}$ . In the later case (bayes), it is an  $(m$  by  $k(k-1)/2)$  matrix where  $m$  is the size of the posterior sample.
- mu.accept.rate** Only when method is "bayes". The MCMC acceptance rate(s) of the long-term average parameter(s).
- sigma.accept.rate** Only when method is "bayes". The MCMC acceptance rate(s) of the short-term variability parameter(s).
- tau.accept.rate** Only when method is "bayes". The MCMC acceptance rate(s) of the timescale(s).
- rho.accept.rate** Only when more than one time series data set are used with method = "bayes". The MCMC acceptance rate(s) of the (cross-) correlation(s).
- data.comb** The combined data set if more than one time series data set are used, and `data1` if only one time series data set is used. This output is only available when method is set to "bayes".

## Author(s)

Zhirui Hu and Hyungsuk Tak

## References

Zhirui Hu and Hyungsuk Tak (2020+), "Modeling Stochastic Variability in Multi-Band Time Series Data," arXiv:2005.08049.

## Examples

```
##### Fitting a univariate damped random walk process

#### Fitting a univariate damped random walk process based on a simulation

n1 <- 20
# the number of observations in the data set

obs.time1 <- cumsum(rgamma(n1, shape = 3, rate = 1))
# the irregularly-spaced observation times

y1 <- rnorm(n1)
# the measurements/observations/magnitudes

measure.error.SD1 <- rgamma(n1, shape = 0.01)
# optional measurement error standard deviations,
# which is typically known in astronomical time series data
# if not known in other applications, set them to zeros, i.e.,
# measure.error.SD1 <- rep(0, n1)

data1 <- cbind(obs.time1, y1, measure.error.SD1)
# combine the single time series data set into an n by 3 matrix

# Note that when astronomical time series data are loaded on R (e.g., read.table, read.csv),
# the digits of the observation times are typically rounded to seven effective digits.
# That means rounding may occur, which produces ties in observation times even though
# the original observation times are not the same.
# In this case, type the following code before loading the data.
# options(digits = 11)

res1.mle <- drw(data1 = data1, n.datasets = 1, method = "mle")
# obtain maximum likelihood estimates of the model parameters and
# assign the result to object "res1.mle"

names(res1.mle)
# to see the maximum likelihood estimates,
# type "res1.mle$mu", "res1.mle$sigma", "res1.mle$tau"

res1.bayes <- drw(data1 = data1, n.datasets = 1, method = "bayes",
                 bayes.n.burn = 10, bayes.n.sample = 10)

# obtain 10 posterior samples of each model parameter and
# save the result to object "res1.bayes"

# names(res1.bayes)
# to work on the posterior sample of each parameter, try
# "res1.bayes$mu.accept.rate", "res1.bayes$sigma.accept.rate", "res1.bayes$tau.accept.rate"
# "hist(res1.bayes$mu)", "mean(res1.bayes$mu)", "sd(res1.bayes$mu)",
# "median(log(res1.bayes$sigma, base = 10))",
# "quantile(log(res1.bayes$tau, base = 10), prob = c(0.025, 0.975))"
```

```
##### Fitting a multivariate damped random walk process based on simulations

n2 <- 10
# the number of observations in the second data set

obs.time2 <- cumsum(rgamma(n2, shape = 3, rate = 1))
# the irregularly-spaced observation times of the second data set

y2 <- rnorm(n2)
# the measurements/observations/magnitudes of the second data set

measure.error.SD2 <- rgamma(n2, shape = 0.01)
# optional measurement error standard deviations of the second data set,
# which is typically known in astronomical time series data
# if not known in other applications, set them to zeros, i.e.,
# measure.error.SD2 <- rep(0, n2)

data2 <- cbind(obs.time2, y2, measure.error.SD2)
# combine the single time series data set into an n by 3 matrix

res2.mle <- drw(data1 = data1, data2 = data2, n.datasets = 2, method = "mle")

# obtain maximum likelihood estimates of the model parameters and
# assign the result to object "res2.mle"

res2.bayes <- drw(data1 = data1, data2 = data2, n.datasets = 2, method = "bayes",
                 bayes.n.burn = 10, bayes.n.sample = 10)

# obtain 10 posterior samples of each model parameter and
# save the result to object "res2.bayes"

# names(res2.bayes)
# to work on the posterior sample of each parameter, try
# "hist(res2.bayes$mu[, 1])", "colMeans(res2.bayes$mu)", "apply(res2.bayes$mu, 2, sd)",
# "hist(log(res2.bayes$sigma[, 2], base = 10))",
# "apply(log(res2.bayes$sigma, base = 10), 2, median)",
# "apply(log(res2.bayes$tau, base = 10), 2, quantile, prob = c(0.025, 0.975))"
```

---

drw.sim

---

*Simulating univariate and multivariate damped random walk processes*


---

## Description

The function `drw.sim` simulates time series data set(s) following either univariate or multivariate damped random walk process.

**Usage**

```
drw.sim(time, n.datasets, measure.error.SD, mu, sigma, tau, rho)
```

**Arguments**

time	A vector containing observation times. Let us use $n$ to denote the length of this vector.
n.datasets	Any positive integer value that denotes the number of time series data sets to be simulated. In simulation, there is no upper limit in the number of time series data sets. Let's use $k$ to denote this number of time series data sets.
measure.error.SD	Optional if measurement error standard deviations are known and available. If one time series data set is simulated, it is a vector of length $n$ containing measurement error standard deviations. If more than one time series data sets are simulated, it is an $n$ by $k$ matrix composed of measurement error standard deviations. If such information is not available, it is automatically set to zeros.
mu	A vector of length $k$ , containing the long-term average parameter(s) of the process.
sigma	A vector of length $k$ , containing the short-term variability parameter(s) (standard deviation) of the process.
tau	A vector of length $k$ , containing the timescale parameter(s) of the process.
rho	Required if more than one time series data sets are simulated ( $k > 1$ ). A vector of length $k(k-1)/2$ , containing the cross-correlation parameters of the process. For example, if $k = 3$ , this is a vector composed of $\rho_{12}, \rho_{13}, \rho_{23}$ . If $k = 5$ , this is a vector composed of $\rho_{12}, \rho_{13}, \rho_{14}, \rho_{15}, \rho_{23}, \rho_{24}, \rho_{25}, \rho_{34}, \rho_{35}, \rho_{45}$ .

**Details**

Given the  $n$  observation times and model parameter values (mu, sigma, tau, rho) possibly with known measurement error standard deviations, this function simulates  $k$  time series data sets.

**Value**

The outcome of `drw.sim` is composed of:

- x** An  $n$  by  $k$  matrix composed of  $k$  simulated time series data each with length  $n$ . That is, each column is corresponding to one simulated time series data set.

**Author(s)**

Zhirui Hu and Hyungsuk Tak

**References**

Zhirui Hu and Hyungsuk Tak (2020+), "Modeling Stochastic Variability in Multi-Band Time Series Data," arXiv:2005.08049.

**Examples**

```
##### Simulating a multivariate damped random walk process

n <- 100
k <- 5
obs.time <- cumsum(rgamma(n, shape = 3, rate = 1))

tau <- 100 + 20 * (1 : 5) #rnorm(k, 0, 5)
sigma <- 0.01 * (1 : 5)
#tau <- c(1 : 5) #rnorm(k, 0, 5)
#sigma <- 0.05 + 0.007 * (0 : 4) #rnorm(k, 0, 0.002)
mu <- 17 + 0.5 * (1 : 5)

rho.m <- matrix(0, k, k)
for(i in 1 : k) {
  for(j in 1 : k) {
    rho.m[i, j] = 1.1^(-abs(i - j))
  }
}

rho <- rho.m[upper.tri(rho.m)]

measure.error.band <- c(0.010, 0.014, 0.018, 0.022, 0.026)
measure.error <- NULL
for(i in 1 : k) {
  measure.error <- cbind(measure.error, rnorm(n, measure.error.band[i], 0.002))
}

x <- drw.sim(time = obs.time, n.datasets = 5, measure.error.SD = measure.error,
            mu = mu, sigma = sigma, tau = tau, rho = rho)

plot(obs.time, x[, 1], xlim = c(min(obs.time), max(obs.time)), ylim = c(17, 20),
     xlab = "time", ylab = "observation")
points(obs.time, x[, 2], col = 2, pch = 2)
points(obs.time, x[, 3], col = 3, pch = 3)
points(obs.time, x[, 4], col = 4, pch = 4)
points(obs.time, x[, 5], col = 5, pch = 5)

##### Simulating a univariate damped random walk process

x <- drw.sim(time = obs.time, n.datasets = 1, measure.error.SD = measure.error[, 1],
            mu = mu[1], sigma = sigma[1], tau = tau[1])
plot(obs.time, x)
```

## Description

The R package **Rdrw** provides a toolbox to fit and simulate univariate and multivariate damped random walk processes, possibly with known measurement error standard deviations via state-space representation. The damped random walk process is also known as an Ornstein-Uhlenbeck process or a continuous-time auto-regressive model with order one, i.e., CAR(1) or CARMA(1, 0). The package **Rdrw** adopts Kalman-filtering to evaluate the resulting likelihood function of the model parameters, leading to a linear complexity in the number of unique observation times. The package provides two functionalities; (i) it fits the model and returns the maximum likelihood estimates or posterior samples of the model parameters; (ii) it simulates time series data following the univariate or multivariate damped random walk process.

## Details

Package:	Rdrw
Type:	Package
Version:	1.0.2
Date:	2020-9-8
License:	GPL-2
Main functions:	<a href="#">drw</a> , <a href="#">drw.sim</a>

## Author(s)

Zhirui Hu and Hyungsuk Tak

## References

Zhirui Hu and Hyungsuk Tak (2020+), "Modeling Stochastic Variability in Multi-Band Time Series Data," arXiv:2005.08049.

# Index

`drw`, [2](#), [10](#)

`drw.sim`, [7](#), [10](#)

`Rdrw`, [9](#)

`Rdrw-package (Rdrw)`, [9](#)