Appendix 2. Climate and PNV reconstructions

This appendix shows the details on how the climate and potential natural vegetation (PNV) reconstructions were obtained.

In addition to the functions described below, you need to install the R package called smpds which can be downloaded from GitHub: <http://github.com/special-uor/smpds>

remotes::install\_github("special-uor/smpds")

## Climate reconstructions

Two functions were created to streamline the process:

#### reconstruct\_climate

This function takes the following parametres:

* .data: data frame with information of the sites/entities for which the reconstructions are required. It should contain at least the following three columns: elevation, latitude and longitude.
* dataset: string of characters to be used as the prefix for the output files (e.g. "smpdsv2").
* path: Path where the reference climate files are stored (check Appendix 1).
* output\_path (*optional*): Path where the output files should be stored.
* ... (*optional*): Additional parametres passed to the function smpds::gwr (including cpus, xy\_buffer, z\_buffer, etc.). Run the following command to get more details about this function, ?smpds::gwr

reconstruct\_climate <- function(.data,
 dataset,
 path,
 output\_path = path,
 ...) {
 if (!file.exists(file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_cld\_", Sys.Date(), ".csv")
 ))) {
 tictoc::tic("CLD")
 .data\_cld <- .data %>%
 smpds::gwr(
 varid = "cld",
 .ref =
 file.path(path, "cru\_ts4.04.1901.2019.cld.dat-clim-1961-1990-int.nc"),
 ...)
 .data\_cld %>%
 readr::write\_excel\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_cld\_", Sys.Date(), ".csv")
 ),
 na = "")
 tictoc::toc()
 } else {
 print("Skipping CLD ...")
 }

 if (!file.exists(file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_pre\_", Sys.Date(), ".csv")
 ))) {
 tictoc::tic("PRE")
 .data\_pre <- .data %>%
 smpds::gwr(
 varid = "pre",
 .ref =
 file.path(path,
 "cru\_ts4.04.1901.2019.pre.dat-clim-1961-1990-int.nc"),
 ...)
 .data\_pre %>%
 readr::write\_excel\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_pre\_", Sys.Date(), ".csv")
 ),
 na = "")
 tictoc::toc()
 } else {
 print("Skipping PRE ...")
 }

 if (!file.exists(file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_tmp\_", Sys.Date(), ".csv")
 ))) {
 tictoc::tic("TMP")
 .data\_tmp <- .data %>%
 smpds::gwr(
 varid = "tmp",
 .ref = file.path(path, "cru\_ts4.04-clim-1961-1990-daily.tmp.nc"),
 ...)
 .data\_tmp %>%
 readr::write\_excel\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_tmp\_", Sys.Date(), ".csv")
 ),
 na = "")
 tictoc::toc()
 } else {
 print("Skipping TMP ...")
 }
}

#### postprocessing

This function takes the following parametres:

* .data: data frame with information of the sites/entities for which the reconstructions are required. It should contain at least the following three columns: elevation, latitude and longitude.
* dataset: string of characters used as the prefix for the output files (e.g. "smpdsv2").
* output\_path: Path where the output files are store.
* CPUS (*optional*): Number of CPUs to be used on the computation of Moisture Index (MI).

postprocessing <- function(.data,
 dataset,
 output\_path,
 CPUS = 1) {
 N\_MAX <- nrow(.data)

 # SPLASH is driven by daily temperature (tmp), precipitation (pre),
 # cloud coverage (cld), and latitude.
 suppressMessages({
 .data\_cld <-
 readr::read\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_cld\_", Sys.Date(), ".csv")
 ),
 n\_max = N\_MAX)

 .data\_cld2 <-
 .data\_cld %>%
 smpds::pivot\_data(varname = "cld")

 # Calculate sunshine fraction from cloud cover
 .data\_sf2 <- .data\_cld %>%
 smpds::pivot\_data(scale = -0.01, add = 1, varname = "sf")

 .data\_pre <-
 readr::read\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_pre\_", Sys.Date(), ".csv")
 ),
 n\_max = N\_MAX)
 .data\_pre2 <- .data\_pre %>%
 smpds::pivot\_data(varname = "pre")

 .data\_tmp <-
 readr::read\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_tmp\_", Sys.Date(), ".csv")
 ),
 n\_max = N\_MAX)
 .data\_tmp2 <- .data\_tmp %>%
 smpds::pivot\_data(varname = "tmp")
 })

 .data\_all\_vars <- .data\_sf2 %>%
 dplyr::left\_join(.data\_pre2) %>%
 dplyr::left\_join(.data\_tmp2) %>%
 dplyr::select(site\_name:elevation, sf, pre, tmp)

 .data\_pre\_map <- .data\_pre %>%
 dplyr::rowwise() %>%
 dplyr::mutate(map = sum(dplyr::c\_across(T1:T365), na.rm = TRUE), .before = T1)

 ## Reconstruct climate variables
 .data\_all\_vars2 <- .data\_all\_vars %>%
 smpds::mi(cpus = CPUS) %>%
 smpds::gdd() %>%
 smpds::mat() %>%
 smpds::mtco() %>%
 smpds::mtwa() %>%
 dplyr::bind\_cols(.data\_pre\_map %>% dplyr::select(map))

 .data\_all\_vars2 %>%
 dplyr::select(-sf, -tmp, -pre) %>%
 readr::write\_excel\_csv(
 file.path(
 output\_path,
 paste0(dataset, "\_climate\_reconstructions\_", Sys.Date(), ".csv")
 ),
 na = ""
 )
}

### Example

Using the following toy dataset, we can obtain climate reconstructions.

`%>%` <- magrittr::`%>%`
smpdsv2\_demo <-
 tibble::tribble(
 ~site\_name, ~entity\_name, ~latitude, ~longitude, ~elevation,
 "Madrid W", "Lopez\_a385", 40.408461, -4.4047361, 550,
 "Lya ljunghed", "Lya ljunghed", 59.585321, 16.727049, 27,
 "Brandenburg", "Matthias\_b17", 52.09969, 14.5223, 126,
 "Dongling Mt", "Dongling Mt 02", 40.03, 115.45, 2280,
 "SB-4, Santa Ana River, Seven Oaks", "SB-4", 34.17735, -116.92446, 1670,
 "Western Siberia, southern taiga, N.Port-Tobolsk 83", "Ryabogina\_a173", 59.478085, 69.141222, 56,
 "Tianshan North", "Tianshan North 48", 43.9581, 88.0833, 1300,
 "Dingbian", "Dingbian 09", 36.920222, 108.172222, 1325,
 "Laguna Carimagua-Bosque", "CABOSQUE\_10", 4.5865, -71.327135, 180,
 "Stand 47 (Mack and Bryant 1974]", "RNM047", 46.9, -120.47, 610
 )

reconstruct\_climate(
 .data = smpdsv2\_demo,
 dataset = "smpdsv2",
 path = "/path/to/CRU/4.04/climatologies",
 output\_path = "/path/to/climate\_reconstructions",
 cpus = 10
)

postprocessing(
 .data = smpdsv2\_demo,
 dataset = "smpdsv2",
 output\_path = "/path/to/climate\_reconstructions",
 cpus = 10
)

#### Output files

"smpdsv2\_climate\_reconstructions\_YYYY-MM-dd.csv"
"smpdsv2\_climate\_reconstructions\_cld\_YYYY-MM-dd.csv"
"smpdsv2\_climate\_reconstructions\_pre\_YYYY-MM-dd.csv"
"smpdsv2\_climate\_reconstructions\_tmp\_YYYY-MM-dd.csv"

#####

##### Demo data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| site\_name | entity\_name | latitude | longitude | elevation |
| Madrid W | Lopez\_a385 | 40.40846 | -4.404736 | 550 |
| Lya ljunghed | Lya ljunghed | 59.58532 | 16.727049 | 27 |
| Brandenburg | Matthias\_b17 | 52.09969 | 14.522300 | 126 |
| Dongling Mt | Dongling Mt 02 | 40.03000 | 115.450000 | 2280 |
| SB-4, Santa Ana River, Seven Oaks | SB-4 | 34.17735 | -116.924460 | 1670 |
| Western Siberia, southern taiga, N.Port-Tobolsk 83 | Ryabogina\_a173 | 59.47809 | 69.141222 | 56 |
| Tianshan North | Tianshan North 48 | 43.95810 | 88.083300 | 1300 |
| Dingbian | Dingbian 09 | 36.92022 | 108.172222 | 1325 |
| Laguna Carimagua-Bosque | CABOSQUE\_10 | 4.58650 | -71.327135 | 180 |
| Stand 47 (Mack and Bryant 1974] | RNM047 | 46.90000 | -120.470000 | 610 |

##### Demo data: climate reconstructions

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| site\_name | entity\_name | latitude | longitude | elevation | mi | gdd0 | mat | mtco | mtwa | map |
| Dingbian | Dingbian 09 | 36.92022 | 108.172222 | 1325 | 0.3949496 | 3592.723 | 8.5509216 | -6.614903 | 21.89719 | 436.0623 |
| Dongling Mt | Dongling Mt 02 | 40.03000 | 115.450000 | 2280 | 0.3909797 | 1573.682 | -2.1796593 | -20.455339 | 13.74946 | 363.9143 |
| Tianshan North | Tianshan North 48 | 43.95810 | 88.083300 | 1300 | 0.1424519 | 3283.318 | 5.4082152 | -13.998738 | 21.72689 | 148.6745 |
| Brandenburg | Matthias\_b17 | 52.09969 | 14.522300 | 126 | 0.8062524 | 3263.467 | 8.7913396 | -1.305118 | 18.06216 | 579.9318 |
| Western Siberia, southern taiga, N.Port-Tobolsk 83 | Ryabogina\_a173 | 59.47809 | 69.141222 | 56 | 1.0226479 | 2046.975 | -0.6118333 | -19.834029 | 18.43383 | 506.5456 |
| Lya ljunghed | Lya ljunghed | 59.58532 | 16.727049 | 27 | 0.9204521 | 2392.763 | 5.6346715 | -4.018724 | 15.94517 | 547.1087 |
| SB-4, Santa Ana River, Seven Oaks | SB-4 | 34.17735 | -116.924460 | 1670 | 0.3944566 | 4309.167 | 11.8059377 | 3.148138 | 23.39599 | 557.0764 |
| Stand 47 (Mack and Bryant 1974] | RNM047 | 46.90000 | -120.470000 | 610 | 0.5790118 | 3472.610 | 9.3301386 | -1.211072 | 19.94392 | 504.7443 |
| Madrid W | Lopez\_a385 | 40.40846 | -4.404736 | 550 | 0.3488190 | 5356.987 | 14.6766758 | 6.367875 | 24.81433 | 439.0107 |
| Laguna Carimagua-Bosque | CABOSQUE\_10 | 4.58650 | -71.327135 | 180 | 1.7895336 | 9651.871 | 26.4434832 | 25.021202 | 27.96895 | 2649.5667 |

##

## Potential Natural Vegetation (PNV) reconstructions

The climate reconstructions are based on the work by (Hengl et al., 2018). The package smpds includes a function to calculate the PNV for each site/entity, based on their location.

This function takes the following parametres:

* .data: data frame with information of the sites/entities for which the PNV values are required. It should contain at least the following two columns: latitude and longitude.
* buffer: the radius of a buffer around each point from which to extract cell values. If the distance between the sampling point and the centre of a cell is less than or equal to the buffer, the cell is included.
* cpus: number of CPUs to be used on the calculation of the PNV values.

Check more details using ?smpds::parallel\_extract\_biome.

`%>%` <- magrittr::`%>%`
smpdsv2\_demo %>%
 smpds::parallel\_extract\_biome(cpus = 2) %>%
 smpds::biome\_name() %>%
 dplyr::rename(PND = description) %>%
 dplyr::select(-colour) %>%
 smpds::pb()

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| site\_name | entity\_name | latitude | longitude | elevation | ID\_BIOME | PNV |
| Dingbian | Dingbian 09 | 36.92022 | 108.172222 | 1325 | 22 | steppe |
| Dongling Mt | Dongling Mt 02 | 40.03000 | 115.450000 | 2280 | 13 | temperate deciduous broadleaf forest |
| Tianshan North | Tianshan North 48 | 43.95810 | 88.083300 | 1300 | 27 | desert |
| Brandenburg | Matthias\_b17 | 52.09969 | 14.522300 | 126 | 9 | cool mixed forest |
| Western Siberia, southern taiga, N.Port-Tobolsk 83 | Ryabogina\_a173 | 59.47809 | 69.141222 | 56 | 15 | cold evergreen needleleaf forest |
| Lya ljunghed | Lya ljunghed | 59.58532 | 16.727049 | 27 | 15 | cold evergreen needleleaf forest |
| SB-4, Santa Ana River, Seven Oaks | SB-4 | 34.17735 | -116.924460 | 1670 | 20 | xerophytic woods/scrub |
| Stand 47 (Mack and Bryant 1974] | RNM047 | 46.90000 | -120.470000 | 610 | 22 | steppe |
| Madrid W | Lopez\_a385 | 40.40846 | -4.404736 | 550 | 4 | warm-temperate evergreen and mixed forest |
| Laguna Carimagua-Bosque | CABOSQUE\_10 | 4.58650 | -71.327135 | 180 | 3 | tropical deciduous broadleaf forest and woodland |

## References

Hengl, T., Walsh, M.G., Sanderman, J., Wheeler, I., Harrison, S.P., Prentice, I.C., 2018. Global mapping of potential natural vegetation: an assessment of machine learning algorithms for estimating land potential. PeerJ 6, e5457. https://doi.org/10.7717/peerj.5457