notation

Notation

Symbol	Meaning
$c_1, c_2, \cdots, c_z$	the thresholds for simplification
$C_0, C_1, \cdots, C_z$	the groups for simplication
$f_\phi(.)$	the function on temporal processing with parameters ϕ
$F_\eta(.)$	the CDF to fit the data with parameter η
$g_\theta(.)$	the function on spatial aggregation with parameter θ
$h(.)$	the function on variable transformation, no parameter
$i : 1, 2, \cdots, n$	the space subscript in the data
$j : 1, 2, \cdots, J$	the temporal subscript in the data
j^*	variation of temporal subscrip in normalising
j'	variation of temporal subscrip in temporal processing
k	time scale in temporal processing
$m : 0, 1, \cdots, 11$	remainder of $j^* \mod 12$ for index monthly data
n	the maximum space index in the data
$p : 1, 2, \cdots, P$	the (maximum) variable index in the data
$p^* : 1, 2, \cdots, P^*$	the dimension reduced variable index
s, $\mathcal{D}_s$	the spatial index, and its space, of the data
t, $\mathcal{D}_t$	the temporal index, and its space, of the data
u	the subtracted constant in benchmarking
α	the subtracted quantity in scaling
γ	the divided quantity in scaling
η, H	parameters, and its space, in PDF distribution fit
θ, Θ	additional parameters, and its space, in spatial aggregation
λ	the PCA factor loadings
σ	the standard deviation of the data in scaling
τ, T	additional parameters, and its space, in variable transformation
$\Phi^{-1}(.)$	the quantile (inverse CDF) function
ψ, Ψ	additional parameters, and its space, in temporal processing

Step	Notation	Notes
Raw data	$\mathbf{x}(\mathbf{s};\mathbf{t})$, $x_p(s_i;t_j)$	$\mathbf{s} \in \mathcal{D}_s$, $\mathcal{D}_s \subseteq \mathbb{R}^2$, $\mathbf{s} = (s_1, s_2, \cdots, s_n)'$ $t \in \mathcal{D}_t$, $\mathbf{t} = (t_1, t_2, \cdots, t_J)'$ $\mathbf{x}(\mathbf{s};\mathbf{t}) = (x_1(\mathbf{s};\mathbf{t}), x_2(\mathbf{s};\mathbf{t}), \cdots, x_P(\mathbf{s};\mathbf{t}))'$ when the pipeline step can be written in univariate case, the data will be referred to as $x(\mathbf{s};\mathbf{t})$
Spatial aggregation	$g_\theta(x(\mathbf{s};\mathbf{t}))$	where $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$, d_θ is the number of parameter of θ
scaling	$\frac{x(s_i;t_j)-\alpha}{\gamma}$	Example 1: z-score standardisation: $\alpha = \bar{x}(s;t)$ and $\gamma = \sigma(s;t)$ to standardise the data by the its mean and standard deviation across all time and all space units. Example 2: standardise into unit interval: $\alpha = \min[x(s_i, t_j)]$ and $\gamma = \max[x(s_i, t_j)] - \min[x(s_i, t_j)]$ where min[.] and max[.] are the minimum and maximum value across all spatial and temporal units.
Normalising	$\Phi^{-1}[F_\eta(x(\mathbf{s};\mathbf{t}))]$	where $\eta \in H \subseteq \mathbb{R}^{d_\eta}$, d_η is the number of parameter of η Example: When CDF $F(.)$ is separately fitted for each month: $\Phi^{-1}[F^m_\eta(x(\mathbf{s};t_{j^*}))]$ where j^* is all the indexes that satisfy $j^* \mod 12 = m$ for each $m = 0, 1, \cdots, 11$
Variable transformation	$h_\tau(x(\mathbf{s};\mathbf{t}))$	where $\tau \in T \subseteq \mathbb{R}^{d_\tau}$, d_τ is the number of parameter of τ TODO: to be filled
Temporal processing	$f_\psi(x(\mathbf{s};\mathbf{t}))$	where $\psi \in \Psi \subseteq \mathbb{R}^{d_\psi}$, d_ψ is the number of parameter of ψ Example 1: aggregate across a time scale of $\psi = k$: $x(s_i;t_{j'}) = \sum_{j=j'-k+1}^{j'} x(s_i;t_j)$ Example 2: aggregate with a kernel weight $\psi = w_{ij}$: $x(s_i;t_{j'}) = \sum_{j=j'-k+1}^{j'} w_{ij} x(s_i;t_j)$

Step	Notation	Notes
Dimension reduction	$x_{p^*}(\mathbf{s};\mathbf{t}) \rightarrow x_p(\mathbf{s};\mathbf{t})$ where $p^* = 1, 2, \cdots, P^*$ and $P^* < P$	Example: first principal component: $x_{p^*}(\mathbf{s};\mathbf{t}) = \sum_{p=1}^{P} \lambda_p x_p(\mathbf{s};\mathbf{t})$ where λ_p is the loading of the PC1, derived from maximising the variance of the data given the constraint $\sum_{p=1}^{P} \lambda_p^2 = 1$
benchmarking	$u[x(s_i, t_j)]$	where u is a scalar of interest in the index constructed, could be a constant or a function of the data, i.e. mean. Example: In SPI, $u = -2$ is the threshold for extreme drought.
Simplification	$\begin{cases} C_0 & c_1 \leq (s_i;t_j) < c_0 \\ C_1 & c_2 \leq x(s_i;t_j) < c_1 \\ C_2 & c_3 \leq x(s_i;t_j) < c_2 \\ ... & \\ C_z & c_z \leq x(s_i;t_j) \end{cases}$	Example: In SPI, four categories are classified: mild drought: $[-0.99, 0]$; moderate drought: $[-1.49, -1]$; severe drought: $[-1.99, -1.5]$, and extreme drought: $[-\infty, -2]$. Here C_0, C_1, C_2, C_3 are the drought categories: mild, moderate, severe, and extreme drought ($z = 3$) and $c_0 = 0, c_1 = -1, c_2 = -1.5, c_3 = -2$
skdjflaksjdflkjsldkfjls	skdjflaksjdflkjsldkfjls	skdjflaksjdflkjsldkfjls