Supplementary material

Seasonal temperature fluctuation and snail adaptive behaviors yield insights into the dynamics and distribution of schistosomiasis in Africa

Mechanistic model formulation

$$\frac{dS}{dt} = (\nu_s(T) - (S+P+I)\nu)(S+rP) - \big(\lambda(T)+\mu(T)\big)S - \varphi_{in}(T)S + \varphi_{out}(T)S_a$$

$$\frac{dS_a}{dt} = \varphi_{in}(T)S - \mu_a(T)S_a - \varphi_{out}(T)S_a$$

$$\frac{dP}{dt} = \lambda(T)S - \big(\sigma_s(T)+\mu_i(T)\big)P - \varphi_{in}(T)P + \varphi_{out}(T)P_a$$

$$\frac{dP_a}{dt} = \varphi_{in}(T)P - \mu_{i_a}(T)P_a - \varphi_{out}(T)P_a$$

$$\frac{dI}{dt} = \sigma_s(T)P - \mu_i(T)I - \varphi_{in}(T)I + \varphi_{out}(T)I_a$$

$$\frac{dI_a}{dt} = \varphi_{in}(T)I - \mu_{i_a}(T)I_a - \varphi_{out}(T)I_a$$

$$\frac{dW}{dt} = \beta_h(T)\,\frac{\nu_c(T)}{\mu_c(T)}I \quad - \sigma_p W$$

$$\frac{dW_m}{dt} = \sigma_p W - \big(\mu_h+\mu_p\big)W_m$$

Where S is susceptible (not exposed to miracidia yet) snails, S_a is susceptible snails, but in the adaptive behavior stage, P is prepatent (exposed to miracidia, but not shed cercaria) snails, P_a is prepatent snails, but in the adaptive behavior stage, I is infected (exposed to miracidia and shed cercaria) snails, I_a is infected snails, but in the adaptive behavior stage, W is the mean number of immature worm in the population, and W_m is the mean number of mature worm in the population. $\lambda(T)$ is the force of infection given by

$$\lambda(T) = \Lambda\left(1 - e^{-\beta_s(T)\frac{M^*}{N}}\right) = \Lambda\left(1 - e^{-\beta_s(T)\frac{\frac{h\delta_e(T)\nu_e W_m}{\mu_m(T)}}{S+P+I}}\right)$$

Table 1: The parameters' description

Parameter	Description	Unit
$\beta_s(T)$	Transmission rate of schistosomiasis in snails	$\frac{s}{p}$

$\beta_h(T)$	Transmission rate of schistosomiasis in humans	$1/h * d$
$v_s(T)$	Fecundity rate of snails	$1/d$
$\mu(T)$	Mortality rate of snails	$1/d$
$\mu_a(T)$	Mortality rate of snails in adaptive behavior stage	$1/d$
$\delta_e(T)$	Probability of hatching success of miracidia	p/e
$\mu_m(T)$	Mortality rate of miracidia	$1/d$
$\mu_c(T)$	Mortality rate of cercariae	$1/d$
$\sigma_s(T)$	Prepatent period in snails	$1/d$
$\mu_i(T)$	Mortality rate of infected snails	$1/d$
$\mu_{i_a}(T)$	Mortality rate of infected snails in adaptive behavior stage	$1/d$
$v_c(T)$	Number of cercariae released by a snail per day	$p/s*d$
$\varphi_{in}(T)$	Transitioning rate into the adaptive behavior stages	$1/d$
$\varphi_{out}(T)$	Transitioning rate out the adaptive behavior stages	$1/d$
v	Factor of density-dependent fecundity rate	$1/s*d$
r	Reduction factor of fecundity rate due to infection	-
v_e	Number of eggs produced by a mature worm	$e/p*d$
Λ	Maximum rate of snail invasion	$1/d$
σ_p	Maturation rate of worms in human body	$1/d$
h	Constant human population having worms	h
μ_p	Mortality rate of worms	$1/d$
μ_h	Mortality rate of worms due to human death	$1/d$

All the details of parameters in the Table 1 can be found in [1] , but the transitioning rates of adaptive behavior stages, which are as follow

$$\varphi_{in}(T) = \frac{e_o}{1 + e^{-\alpha(T-T_o)}}$$

$$\varphi_{out}(T) = \frac{e_o}{1 + e^{\alpha(T - T_o)}}$$

Where the parameter T_o is the temperature at which $\varphi_{in}(T)$ and $\varphi_{out}(T)$ reach half of their maximum saturation value e_o, and α is inversely related to the steepness of the transitioning rates.

Converting Mean Parasite Burden to Prevalence

Note that our model simulations produce the mean parasites burden (MPB) for the population. We convert this value into the prevalence of schistosomiasis in the population. We assume the distribution of parasites in the population is following a negative binomial distribution. By the formula obtained in [2,3] we convert MPB into a prevalence.

$$P(m,\ k) = 1 - \left(1 + \frac{m}{k}\right)^{-k}$$

Where $P(m, k)$ is the probability of having one or more parasites, m = MPB (Mean Parasite Burden) and k is dispersion (aggregation) parameter which is estimated by leveraging the data of prevalence and the mean number of eggs collected in the stool and urine from 16 villages from 2016 to 2018 in the lower basis of Senegal river. We separately estimated m and k for each village via maximum-likelihood estimation by using the prevalence in the data. We found a significant correlation ($p << 0.05$) between log transformed estimated k and m after fitting a linear regression ($k = \frac{m^{0.615}}{127}$ for *S. mansoni* and $k = \frac{m^{0.519}}{26}$ for *S. haematobium*). Thus, we used this relationship of regression to calculate the aggregation parameter for a given m. Therefore, for MPB of our simulations, we found k and later the prevalence of schistosomiasis $P(m,\ k)$.

Derivation of seasonality metric

We assume the quantity of seasonality can be drawn from the distance between the mean temperature of the warmest quarter (Bio10) and the mean temperature of the coldest quarter (Bio11). If we assume the temperature time series can be modeled with a sinusoidal function below

$$T(t) = T_{med}\left(1 + \varepsilon \sin\left(\frac{2\pi t}{365}\right)\right)$$

Then, we can calculate the average of the warmest and coldest quarters for this temperature time series. If a year is 365 days, then the warmest quarter of season occurs in

$$t \in \left[\frac{365}{8}, \frac{3 * 365}{8}\right]$$

Therefore, the average temperature of warmest quarter

$$\frac{1}{\frac{3*365}{8}-\frac{365}{8}}\int_{\frac{365}{8}}^{\frac{3*365}{8}} T_{med}\left(1+\epsilon\, sin\left(\frac{2\pi t}{365}\right)\right)dt$$

$$=\frac{8}{2*365}T_{med}\left(t-\epsilon\, cos\left(\frac{2\pi t}{365}\right)\frac{365}{2\pi}\right|_{\frac{365}{8}}^{\frac{3*365}{8}}$$

$$=T_{med}+T_{med}\epsilon\frac{2\sqrt{2}}{\pi}$$

Similarly, the average temperature of coldest quarter

$$t\in\left[\frac{5*365}{8},\frac{7*365}{8}\right]$$

and with the same calculus,

$$T_{med}-T_{med}\epsilon\frac{2\sqrt{2}}{\pi}$$

The difference between warmest and coldest quarters is

$$T_{med}\epsilon\frac{4\sqrt{2}}{\pi}$$

This value represents the seasonality value for us as long as we use sinusoidal function for the temperature time series. If we want to draw the ϵ for a given location with Bio01, Bio10, Bio11, then

$$T_{med}\epsilon\frac{4\sqrt{2}}{\pi}=Bio10\text{-}Bio11\Rightarrow\epsilon=\frac{(Bio10\text{-}Bio11)\pi}{T_{med}4\sqrt{2}}=\frac{(Bio10\text{-}Bio11)\pi}{(Bio01)4\sqrt{2}} \qquad (1)$$

Parameter estimation for the adaptive behavior function and model comparison

First, we created two-dimensional grid with ε and T_{med} where $\varepsilon\in[0.02,0.26]$ with increasing step 0.02 and $T_{med}\in[14,31]$ with increasing step 0.5. Then, we calculated seasonality value ε for each GNTD data by using Bio01, Bio10 and Bio11 with formula given in (1). Then, we assigned each GNTD data into two-dimensional ε and T_{med} grid based on its value of ε and T_{med} (Bio01). Therefore, each grid cell of ε and T_{med} might have several GNTD data, or NAN. If there are more than one GNTD data in a grid cell then,

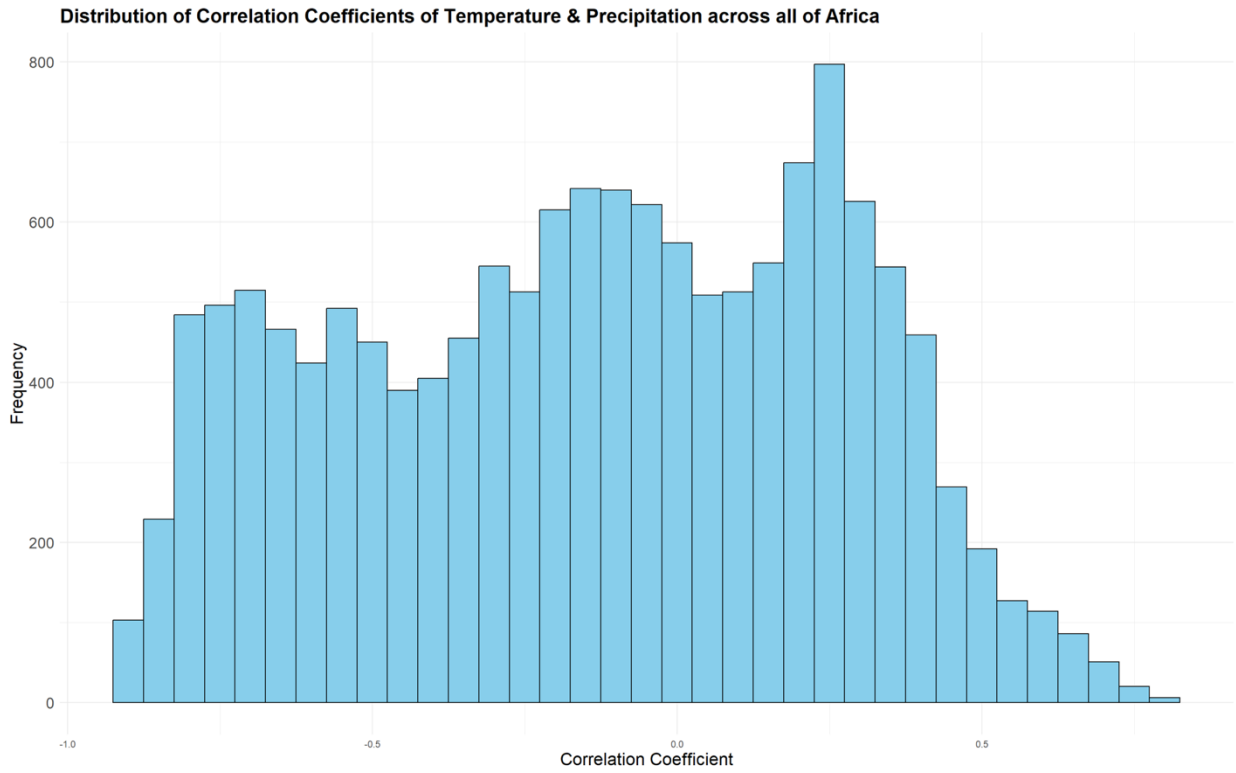
we sum up the total number people tested and positive cases in this each grid cell to obtain a single data point for the grid cell. This this two-dimensional grid will represent our data set now. The reason we do this, because we do not want to run the model for each GNTD data (> 11,000 points) for the sake of computation cost while most data point have almost same ε and T_{med} (Bio01) values. On the other hand, we created three-dimensional grid with the set of ($e_0,\ T_0,\ \alpha$) vectors where $e_o \in [0.005, 0.1]$ with increasing step 0.02, $T_o \in [22.5, 27]$ with increasing step 0.5, and $\alpha \in [0.1, 1.5]$ with increasing step 0.2. We ran the model of adaptive behavior for each of set of ($e_0,\ T_0,\ \alpha$) with temperature time series

$$T(t) = T_{med}\left(1 + \varepsilon \sin\left(\frac{2\pi t}{365}\right)\right)$$

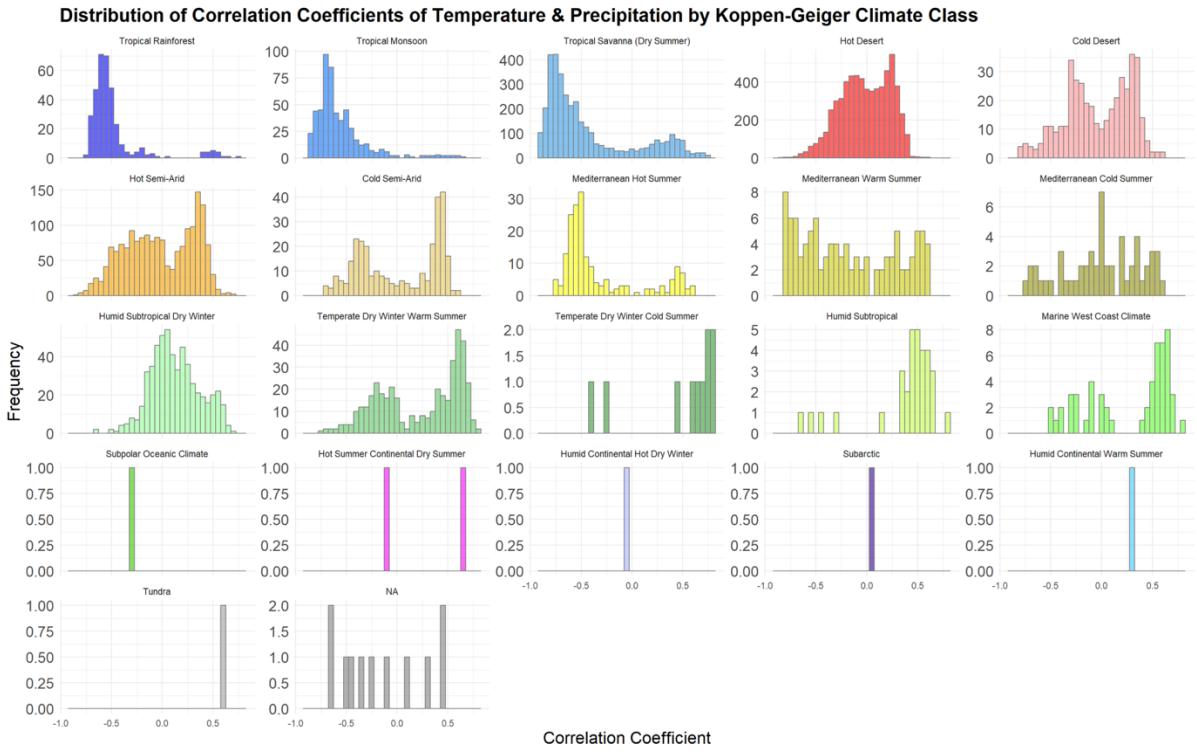
whose ε and T_{med} are same as two-dimensional grid previously describe, $\varepsilon \in [0.02, 0.26]$ with increasing step 0.02 and $T_{med} \in [14, 31]$ with increasing step 0.5. Therefore, we obtained the two-dimensional simulation of prevalence matrix for each of value of ($e_0,\ T_0,\ \alpha$). We obtained maximum likelihood value for each ($e_0,\ T_0,\ \alpha$) By assuming that the distribution of schistosomiasis in each cell is binomial distributed with the probability of success to be the prevalence value of simulations, the number of trails to be total number of cases and the number of successes is the positive case. Finally, we choose the set of parameters ($e_0,\ T_0,\ \alpha$) which maximize the likelihood function. We found that the parameter set (23, 0.065, 0.7) for *S. heamatobium* and (23, 0.085, 0.9) for *S. mansoni* maximize the likelihood function. We also calculated the maximum likelihood value for the model with seasonal temperature without adaptive behavior and model with constant temperature. Based on the AIC (2*ln (maximum likelihood value) + 2*(number of independent parameters)) value we compared three models' performance. We found AIC values for *S. heamatobium*: 212050.7 for the model with aestivation, 673675.8 for the model with seasonal temperature and without aestivation and 1502795 for the model with constant temperature. Similarly, 218688.6 for the model with aestivation, 920650.5 for the model with seasonal temperature and without aestivation, and 876934.2 for the model with constant temperature.

Temperature can be a proxy for aestivations

We draw a 30-year time series of mean monthly temperature and precipitation data from the ERA5 dataset across the entirety of Africa. To account for spatial autocorrelation, we aggregated the data within 10×10 km (100 km²) quadrats, regularly spaced 50 km apart across the continent and, then we computed the correlation between the 360 monthly temperature and precipitation values at each of the 14,596 unique locations.

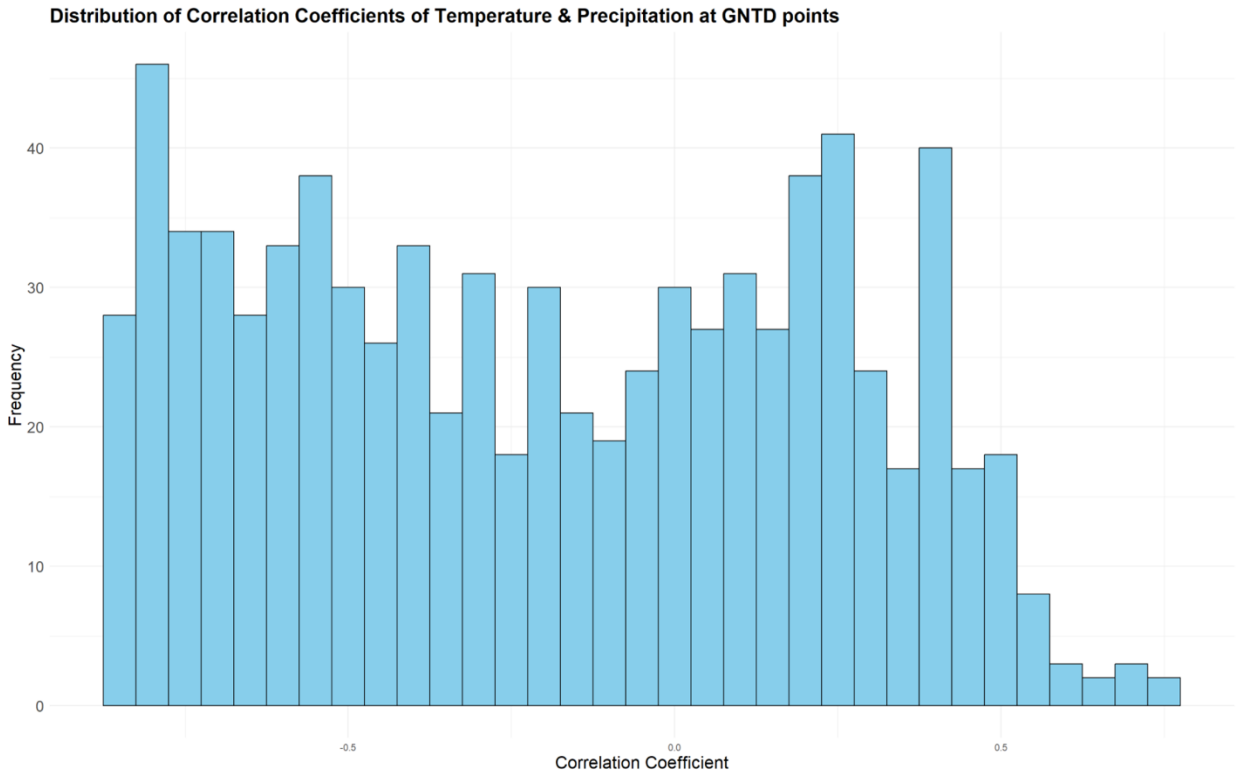


Less than 5% of the continent exhibits a temperature-precipitation correlation exceeding +50%. In contrast, approximately two-thirds of the regions show a negative correlation between these variables. Additionally, we calculated the correlation within the Köppen-Geiger climatic classifications, with the results presented in the figure below.

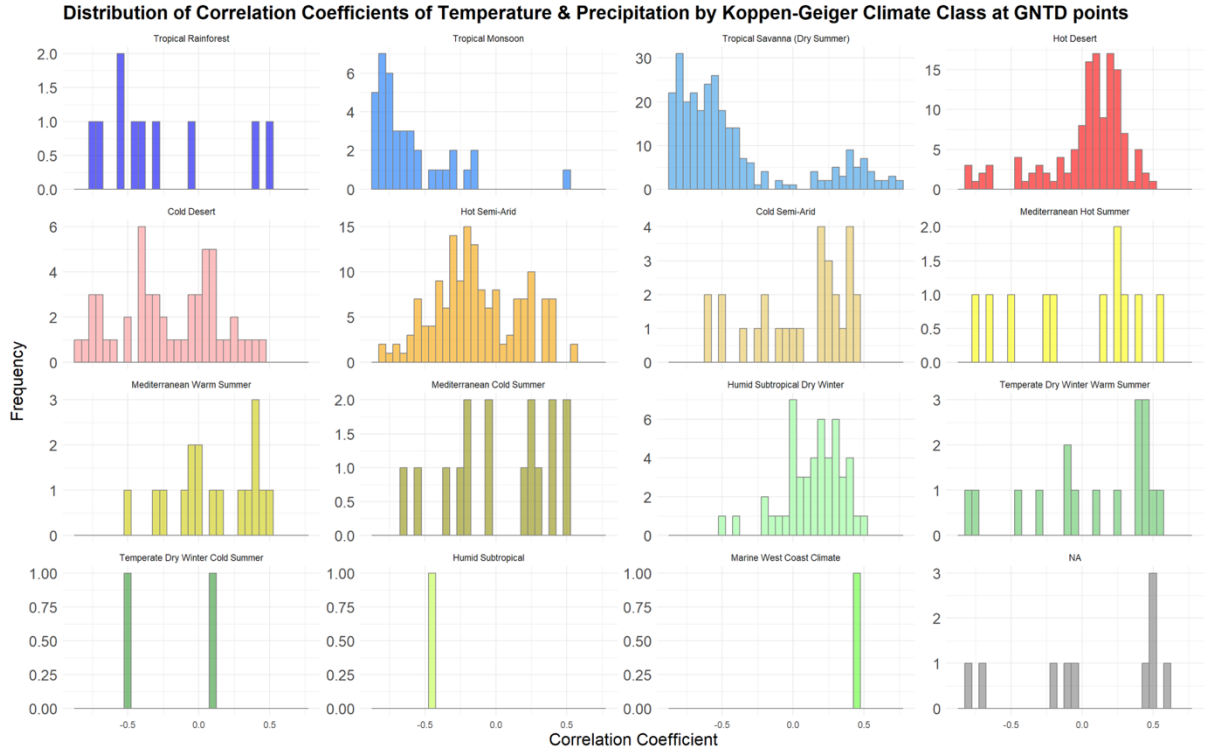


Several of these distributions exhibit bimodality, with the right mode indicating a positive correlation between temperature and precipitation. This pattern is particularly evident in climatic classes such as “Temperate Dry-Winter, Warm-Summer,” “Hot Desert,” and “Hot Semi-Arid.” However, most areas display either a weak or negative correlation.

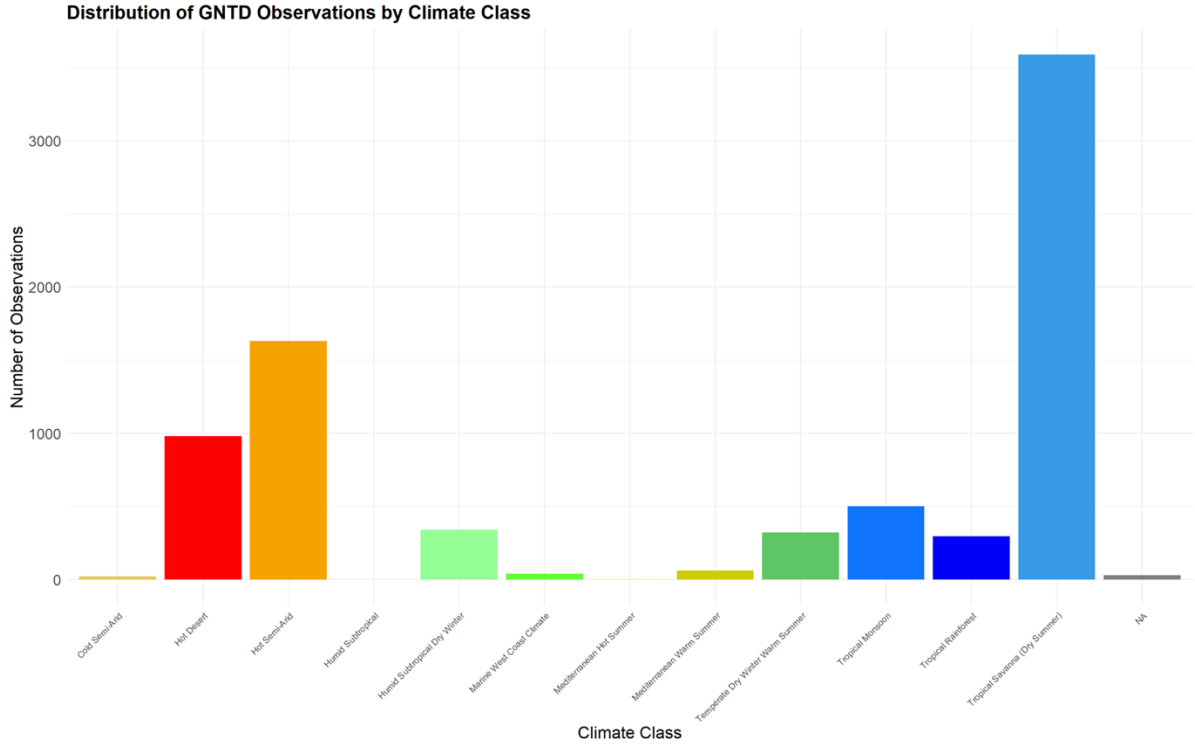
Additionally, we repeated the analysis by resampling ERA5 data over a 100 km × 100 km grid, considering only grid cells that contained at least one GNTD sampling point. This approach yielded 822 unique locations—significantly fewer than the >14,000 locations used to derive the distributions above. Nevertheless, the results remained consistent, as presented below.



This is substantially true also when we look at distribution for each climatic class, as reported here below:



Interestingly, the majority of GNTD data points are located in dry tropical savanna regions (see histogram below), where temperature and precipitation predominantly exhibit a negative correlation.



Our findings support the assumption that, in general, temperature serves as a valid proxy for thermal stress in snails, with a few notable exceptions mentioned earlier. This includes its role in driving aestivation, particularly in most locations where temperature and precipitation are negatively correlated.

As a side note, we are also analyzing HOBO water temperature data from 17 unique sampling locations in Kenya, including rivers, streams, and Lake Victoria, collected from Sam Locker's sites. Interestingly, at nearly all locations, the mean daily water temperature is equal to or even higher than the corresponding ERA5 air temperature. Although our field sites in Senegal provide only a limited number of long-term water temperature records, our findings there are largely consistent with these results.

In both Kenya and Senegal, snail vectors have been observed at considerable depths, ranging from 3–5 meters below the water surface to as deep as 20 meters. However, the extent to which these deep-water snails contribute to disease transmission remains unclear. Prakash's experiments have shown that cercariae are negatively buoyant, meaning they sink unless actively swimming. Additionally, deep water bodies are typically stratified, with a thermocline occurring between 2 and 5 meters below the surface.

Given these factors, it is reasonable to argue that snails residing below the thermocline are unlikely to play an active role in transmission. Cercariae emerging from these depths would have to swim several meters from colder, deeper waters through the thermocline to reach the surface—a highly demanding effort. While further studies will explore this aspect in greater

detail, it is plausible to assume that this behavioral adaptation, while potentially extending snail lifespan, may temporarily disrupt transmission.