Generalized feature vectors (GFV) are denoted generally by the ordered pair (a, b), which in this context can be read as "a spans of b days", where a span is simply a length of time. For example, a week is one span of seven days, so $(1, 7)$, three weeks is $(3, 7)$, one month is $(1, 30)$, thirty days is $(30, 1)$, etc. A GFV is then formed by averaging the normalized data of the b days for each of the a spans. An example of this process is given in Figure 2 of the main article. Once a GFV is decided, the target feature vector $\vec{t}$ can be built. For the sake of intuitive correlation, if t is a day, $\vec{t}$ is its feature vector. Take note of the ab days prior to and including t, as well as their daily values in each ordered set for those days from $P, T_{min} and T_{max}$. Let p_d^τ denote the datum d days prior to day τ (used as a placeholder for t or h_{ir}) from the ordered set p. In order to build $\vec{t}$, one must first build vectors $P^{\vec{t}}, T_{min}^{\vec{t}}$, and $T_{max}^{\vec{t}}$, where

$$p_y^{\vec{t}} = \frac{1}{b} \sum_{j=0}^{b-1} p_{j+by}^t \text{ for } p = P, T_{min}, T_{max}, \text{ for } y = 0, ..., a-1 \tag{S1}$$

and

$$\vec{t} = [P^{\vec{t}}, T_{min}^{\vec{t}}, T_{max}^{\vec{t}}]$$

to ensure all entries are compartmentalized by the variable they represent. This feature vector will include the data from the target date, and likewise the historical feature vectors will include the data from their respective analogue dates.

Now that the target feature vector has been found, the potential analogues that should be found can be identified and their feature vectors can be calculated and built. One way to go about choosing potential analogues would be to simply use the date of the target day for all available years. For example, if t was April 15th 2022, only the time from April 15th 2021 to whatever the earliest year of available data would be considered. However, as the existence of leap years indicates, years do not run in perfect cycles. It is entirely possible a better analogue could occur on April $14^{th}, 16^{th}$, or even 8th. To account for this, a fifteen-day temporal window centred around April 15th was used in the study. This means that in addition to April 15th, the seven days before and after it —meaning April 8th to April 22nd— in each year were considered as potential analogues. With this in mind, let h_{ir} be the ith day of the temporal window of the rth year for $i = 1, ..., 15$ for $r = 1, ..., n-1$, where n is the number of years, including the incomplete current year. Then, its feature vector, $\vec{h}_{ir}$, is

$$\vec{h}_{ir} = [P^{\vec{h}_{ir}}, T_{min}^{\vec{h}_{ir}}, T_{max}^{\vec{h}_{ir}}]$$

with

$$p_y^{\vec{h}_{ir}} = \frac{1}{b} \sum_{j=0}^{b-1} p_{j+by}^{h_{ir}} \text{ for } p = P, T_{min}, T_{max},$$
$$\text{for } y = 0, ..., a-1. \text{ for } i = 1, ..., 15, \text{ for } r = 1, ..., n-1. \tag{S2}$$

Now that all the target and historical feature vectors have been identified, they can be compared to determine which is the analogue. With some weight vector

W , defined by the user (for the purposes of this paper all entries of W were set to 1), the Euclidean distance between a target and a historical feature vector is

$$e_{ir} = \sqrt{\sum_{m=1}^{av} W_m \left(\vec{t}_m - (\vec{h}_{ir})_m\right)^2} \text{ for } i = 1, ..., 15, \text{ for } r = 1, ..., n-1 \qquad \text{(S3)}$$

where v is the number of parameters used to build feature vectors —3 in this case— and a is the number of spans from the aforementioned (a, b). Once all Euclidean distances are found, they are sorted, and the days corresponding to the k lowest distances are noted. Let k_{jf} be the prediction generated by the jth nearest neighbour, with the said prediction being the total precipitation for f days after that day. Then let E_j be the Euclidean distance of said prediction's feature vector from that of the target, and $E = \sum_{j=1}^{k} E_j$. Let W_j be the weights — not to be confused with the weight vector w— then makeing the final prediction using kNN be as follows:

$$K_f = \sum_{j=1}^{k} \frac{E_j}{E} k_{jf}. \qquad \text{(S4)}$$