Supplementary Material

Support vector regression

Considering the regression problem, for a given training dataset D = $\{(\boldsymbol{x_1}, \boldsymbol{y_1}), (\boldsymbol{x_2}, \boldsymbol{y_2}), \cdots\cdots, (\boldsymbol{x_m}, \boldsymbol{y_m})\}$, $\boldsymbol{x}_i$ represents a vector of input meteorological data and crop coefficient and $\boldsymbol{y}_i$ denotes the ET_c value calculated by FAO-PM in the training period, we seek to find a function $f(\boldsymbol{x})$, where $f(\boldsymbol{x}) = \omega^T\boldsymbol{x} + \boldsymbol{b}$, ω^T is the weights vector norm and $\boldsymbol{b}$ is a bias, such that the deviation between $f(\boldsymbol{x})$ and y is minimized. Different from traditional regression models, the SVR assumes that we can tolerate a deviation of at most ε between $f(\boldsymbol{x})$ and y. As shown in the figure (Figure S.1), this is equivalent to constructing an interval band of width 2ε centered on $f(\boldsymbol{x})$. If the training sample $(\boldsymbol{x_i}, \boldsymbol{y_i})$ fall into the subinterval band, then the model output $f(\boldsymbol{x})$ is considered to have correctly predicted training sample. The SVR problem can then be expressed as following:

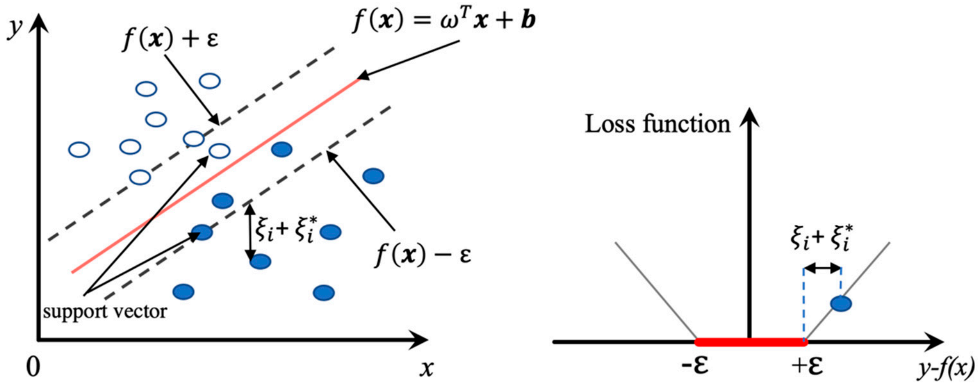
$$\min_{\omega, b, \xi_i, \xi_i^*} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m}(\xi_i + \xi_i^*) \tag{S1}$$

$$subject\ to: f(\boldsymbol{x}_i) - y_i \le \varepsilon + \xi_i,$$

$$y_i - f(\boldsymbol{x}_i) \le \varepsilon + \xi_i,$$

$$\xi_i \ge 0, \xi_i^* \ge 0, i = 1,2,\cdots, m.$$

where C is the regularization constant, b is a bias, ω is the weights vector norm. The regularization constant determines the balance between the value of tolerable deviations greater than ε and the complexity of the function. The ε-insensitive loss function [1] represents the discrepancy between the actual calculated values (FAO-PM-ET_c) and the estimated values (SVR-ET_c). The loss function can be described by introducing (non-negative) slack variables ξ_i and ξ_i^* to measure the deviation of training samples outside the ε-insensitive zone [2] (Figure S.1).



Supplement figure S1. The left figure is a schematic representation of SVR, and the right figure is the ε-insensitive loss function for the SVR model [3]. The circles represent the training samples and the dotted lines represent the discrepancy between the actual calculated values (PM-FAO56-ET_c) and the estimated values (SVR-ET_c).

By introducing Lagrange multipliers (α_i, α_i^*) and making their $(\omega, b, \xi_i, \xi_i^*)$ partial derivatives zero the dual problem of SVR can be obtained:

$$\max_{\alpha_i, \alpha_i^*} \sum_{i=1}^{m} y_i(\alpha_i^* - \alpha_i) - \varepsilon(\alpha_i^* + \alpha_i) - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\,\boldsymbol{x}_i^{T*}\boldsymbol{x}_j \tag{S2}$$

$$subject\ to\colon \sum_{j=1}^{m}(\alpha_i^* - \alpha_i) = 0,$$

$$0 \leq \alpha_i, \alpha_i^* \leq C$$

and the solution to the SVR problem can be transformed into the following expression:

$$f(\boldsymbol{x}) = \sum_{i=1}^{m}(\alpha_i^* - \alpha_i)\boldsymbol{x}_i^T\boldsymbol{x} + b \tag{S3}$$

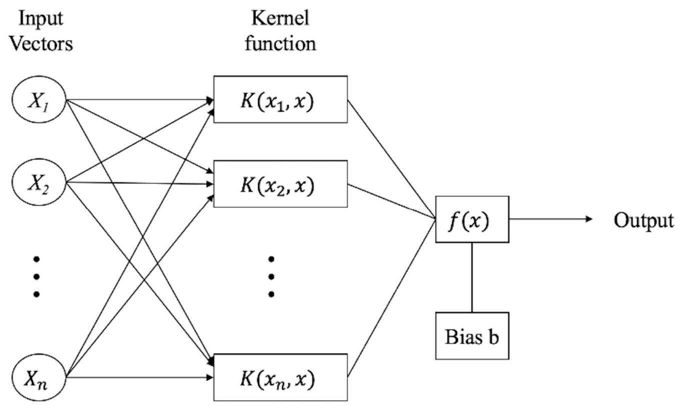
The kernel function allows linearly indistinguishable samples to be mapped from the original space to a higher dimensional feature space [4]. This feature space renders the samples linearly distinguishable, facilitating the application of SVR expressed as,

$$f(\boldsymbol{x}) = \sum_{i=1}^{m}(\alpha_i^* - \alpha_i)\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) + b \tag{S4}$$

where the $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the kernel function. As mentioned before, the choice of the kernel is crucial to the construction of the SVR model, and the performance of different kernels varies. The radial basis function (RBF) kernel is defined as following:

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = exp(-\frac{\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2}{2\gamma^2}) \tag{S5}$$

The SVR model's hyper-parameters include C and RBF parameter γ and ε.The structer of the SVR model is shown in Figure S.2.



Supplement figure S2. Support vector regression structure [3].

Particle swarm optimize algorithm

Assuming that there is a community composed of n particles in the D-dimensional search space, the ith particle in t iterations can be represented by a D-dimensional vector: $X_i^t = (x_{i1}^t, x_{i2}^t, \cdots, x_{iD}^t), (i = 1,2,\cdots,n)$ and the velocity of the i-th particle is $V_i^t = (v_{i1}^t, v_{i2}^t, \cdots, v_{iD}^t), (i = 1,2,\cdots,n)$. The optimal solution found by the ith particle itself is called individual optimum value P_{best}, and the optimal solution found by the whole particle swarm is called the global optimum G_{best}, denoted as following:

$$P_{best}^t = (p_{i1}^t, p_{i2}^t, \cdots, p_{iD}^t), (i = 1,2,\cdots,n) \quad \text{(S6)}$$

$$G_{best}^t = (P_{g1}^t, P_{g2}^t, \cdots, P_{gD}^t) \quad \text{(S7)}$$

The particle positions and velocities for the t+1 iteration are then updated as following:

$$V_i^{t+1} = \omega^t V_i^t + c_1\tau_1(P_{best}^t - X_i^t) + c_2\tau_2(G_{best}^t - X_i^t) \quad \text{(S8)}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \quad \text{(S9)}$$

where c_1 and c_2 are the learning factors, τ_1 and τ_2 are uniform random numbers in the range of [0,1], and ω is the inertia weight.

References

1. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Vapnik, V.N., Ed.; Springer: 1995.
2. Shrestha, N.K.; Shukla, S. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. *Agricultural and Forest Meteorology* **2015**, *200*, 172-184, doi:10.1016/j.agrformet.2014.09.025.
3. Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In Proceedings of the 10th Annual Conference on Neural Information Processing Systems (NIPS), Denver, Co, 1996; pp. 155-161.
4. Chen, S.; He, C.; Huang, Z.; Xu, X.; Jiang, T.; He, Z.; Liu, J.; Su, B.; Feng, H.; Yu, Q.; et al. Using support vector machine to deal with the missing of solar radiation data in daily reference evapotranspiration estimation in China. *Agricultural and Forest Meteorology* **2022**, *316*, doi:10.1016/j.agrformet.2022.108864.