Correlation idea

This is a problem I am encountering in a research project. I have thought of a naive approach to solve it which relies on many assumptions, is not guaranteed to work, and would be tedious and computationally expensive to implement. Hence, I am wondering if there would be cleverer or more general approaches. Below I will try to explain my problem as clearly as possible.

Tree growth at day d is thought of as a function of meteorology that has happened in the past. For instance, tree growth on a given day d may depend, on the cumulative temperature between day *d-1* and day *d-20* but probably not on the temperature a year before. Likewise for all the other meteorological variables.

I have daily meteorological data and daily tree growth data measured in number of cells created.

For any given meteorological variable, I would like to know for which period in the past the variable influences tree growth the most (in terms of daily creation of cells).

Naive approach:

Let's choose any meteorological variable, like temperature.

I have daily mean temperature data $T(d)$ and daily cell creation data $C(d)$. These are discrete functions but I can extend them to all time by interpolation.

Solving the problem is equivalent to finding values $0 < \lambda \leq \mu$ which maximise the correlation

$$K(\mu, \lambda) = cor(\; \frac{1}{2\lambda} \int_{(d-\mu)-\lambda}^{(d-\mu)+\lambda} T(t)\, dt, \quad C(d) \;)$$

2λ would then be length of the time interval which most influences growth and μ the number of days in the past at the center of the interval.

We can refine this further. There is probably a sweet spot in the past for which temperature particularly influences present growth. Conversely, temperatures further away in time from that sweet spot influence present growth less. To take this into account we can weigh the integration with a normal distribution.

Solving the problem then becomes maximising

$$\tilde{K}(\mu, \sigma, \lambda_1, \lambda_2) = cor(\; \frac{\int_{(d-\mu)-\lambda_1}^{(d-\mu)+\lambda_2} T(t)\; F_{d-\mu,\sigma}(t)\; dt}{\int_{(d-\mu)-\lambda_1}^{(d-\mu)+\lambda_2} F_{d-\mu,\sigma}(t)\; dt}, \; C(d) \;)$$

where $F_{d-\mu,\sigma}$ is the density function of a normal distribution of mean $d - \mu$ and standard deviation σ.

Lastly, I am coding in R and the correlation methods available are "Pearson" (the usual one), "Spearman" and "Kendall". Pearson mostly measures linear relationships. However, I would not expect growth to be linearly dependent with the temperature. There could be a polynomial relation with optimum temperatures causing maximum growth and temperatures far from that optimum causing less or no growth. For that reason I was going to use Spearman correlation which better captures non-linear relationships.

With the above approach I need to assume that the meteorological influences (e.g. temperature and rainfall) are not dependent, but they probably are. I am treating each meteorological factor separately for now.

Going back to questions:

- Do you have any thoughts on how I could tackle this problem?
- Have you spotted significant flaws in this method that could be addressed?

- Do you know of better methods to solve this problem?
- Do you know of similar problems that have already been solved, which I could look into for a solution?

Thanks a lot for your help!