Statistical Reference Method

cyrilvoyant

May 2021

1 Algorithms

The pseudo-codes detailed here are adapted to the case of global irradiation ($I_{GH}(t)$ with $t \in \{\overbrace{1, ..., T^*}^{InSample}, \overbrace{T^*, ..., T}^{OutSample}\}$) though they can be modified for any kind of time series. Up to now, we deliberately neglect the phenomena of over irradiance ($I_{GH}(t) \in [0, I_{CS}(t)]$), however depending on the time step, the clear sky model used and the quality of the time-stamp, it could be necessary to multiply I_{CS} by an arbitrary coefficient β (generally between 1 and 2). Anyway, all models must make it possible to provide forecasts for all hours of the day and night, however the validation of the results is only performed on the daytime hours (authorizing solar elevation greater than $5 - 10°$). Even if it is not the purpose of this study, it is important not to neglect the forecasts of the first and last hours of the daylight, they can be very important for energy management systems. Often, the real reasons for which a filtration is operated because of the poor quality of the detection concerning these hours and the strong repercussions (periodic peaks on κ) that a poor time-stamp can induce.

1.1 Persistence

This persistence predictor (Algorithm 1) is certainly the simplest method use in order to operate predictions with reliability.

Algorithm 1 PER

Require: $I_{CS}, I_{GH}, h > 0, \beta \in [1, 2]$
Ensure: $\widehat{I}_{GH}(t + h)$ with $t \in [T^*, T]$

```
n ← 0
repeat
    n ← n + 1
until I_CS(t − n) ≠ 0
Pred← min(I_GH(t − n) × I_CS(t + h)/I_CS(t − n), β × I_CS(t + h))
return Pred
```

1.2 Climatology

Even if this predictor (Algorithm 2) is never used in practice, it is an important way to gauge results in solar prediction study. When no model of knowledge is available, a moving average can be a good alternative. One of the characteristics of this model is that the observed forecast error is constant whatever the horizon considered.

Algorithm 2 CLIM

Require: $I_{CS}, h > 0$
Ensure: $\widehat{I}_{GH}(t+h)$ with $t \in [T^*, T]$
  Pred$\leftarrow I_{CS}(t+h)$
  return Pred

1.3 Climatology Persistence

CLIPER is undoubtedly the new standard of reference forecast for solar irradiation. A filtering process is operated in Algorithm 3. ϵ is usually taken close to 10 (Wh/m^2) while certain authors prefer use a threshold between 5° and 10° concerning the solar elevation. The information linked to the cloudiness being observable only in the presence of daylight, only these moments must be used. This means that at sun-up, it is the data from the day before that is used, so we understand the limit of statistical forecast models using only endogenous quantities.

Algorithm 3 CLIPER

Require: $I_{CS}, I_{GH}, h > 0, \beta \in [1, 2], \epsilon \in [1, 30]$
Ensure: $\widehat{I}_{GH}(t+h)$ with $t \in [T^*, T]$
  for $n := 1$ to T^* **do**
    if $I_{CS}(n) < \epsilon$ **then**
      $\kappa(n) = \emptyset$
    else
      $\kappa(n) \leftarrow I_{GH}(n)/I_{CS}(n)$
    end if
  end for
  $\bar{\kappa} \leftarrow$mean$(\kappa(n))$
  $\rho \leftarrow$ACF$(\kappa(n), \kappa(n-h))$
  $nn \leftarrow 0$
  repeat
    $nn \leftarrow nn + 1$
  until $I_{CS}(t - nn) \geq \epsilon$
  Pred$\kappa \leftarrow min(\rho \times \kappa(t - nn) + (1 - \rho) \times \bar{\kappa}, \beta)$
  Pred $\leftarrow$ Pred$\kappa \times I_{CS}(t+h)$
  return Pred

1.4 Exponential Smoothing

In practice, it is not necessary to calculate the smoothing on all the in-sample data, limiting to a range covering the daily periodicity ($max = 24$ h) or 2 times this ($max = 48$ h) is sufficient to obtain good results (Algorithm 4).

Algorithm 4 ES

Require: $I_{CS}, I_{GH}, h > 0, \beta \in [1, 2], \epsilon \in [1, 30], max \in [10 - 48]$
Ensure: $\widehat{I}_{GH}(t + h)$ with $t \in [T^*, T]$

```
for n := 1 to T* do
    if I_CS(n) < ε then
        κ(n) = 1
    else
        κ(n) ← I_GH(n)/I_CS(n)
    end if
end for
κ̄ ←mean(κ(n))
ρ ←ACF(κ(n), κ(n − h))
for nn := 0 to max − 1 do
    if I_CS(t − nn) < ε then
        κ(t − nn) = 1
    else
        κ(t − nn) ← I_GH(t − nn)/I_CS(t − nn)
    end if
end for
Predκ ← min(ρ × Σ_{i=0}^{max−1} (1 − ρ)^i × κ(t − i) + κ̄ × (1 − ρ)^max, β)
Pred ← Predκ × I_CS(t + h)
return Pred
```

1.5 ARTU

In this version of the code (Algorithm 5), we propose to associate the night hours with a κ equal to 1 but another way which is slightly more complex but which gives very good results consists in neglecting the night hours by removing them completely as done in the Algorithm 3 (**if** $I_{CS}(n) < \epsilon$ **then** $\kappa(n) = \emptyset$). The method requires knowledge of α and K, which is achieved by interpolating the $\mathcal{M}(R)$ matrices (see attached code). Knowing the correlation coefficients ($\rho(h)$ and$\rho(2h)$) and the measurement reliability ($R = 0, 0.01, 0.05, 0.1$) the interpolation allows an estimate of α and K for these three characteristic values.

Algorithm 5 ARTU

Require: $I_{CS}, I_{GH}, h > 0, \beta \in [1, 2], \epsilon \in [1, 30], R \in [0, 0.01, 0.05, 0.1], \mathcal{M}(R)$
Ensure: $\widehat{I}_{GH}(t + h)$ with $t \in [T^*, T]$

```
for n := 1 to T* do
    if I_CS(n) < ε then
        κ(n) = 1
    else
        κ(n) ← I_GH(n)/I_CS(n)
    end if
end for
κ̄ ←mean(κ(n))
ρ1 ←ACF(κ(n), κ(n − h))
ρ2 ←ACF(κ(n), κ(n − 2h))
(α, K) ← interpolate(M(R), ρ1, ρ2, R)
S ← α + K
P ← α × K
for nn := 0 to h do
    if I_CS(t − nn) < ε then
        κ(t − nn) = 1
    else
        κ(t − nn) ← I_GH(t − nn)/I_CS(t − nn)
    end if
end for
Predκ ← min(S × κ(t) − P × κ(t − h) + (1 + P − S) × κ̄, β)
Pred ← Predκ × I_CS(t + h)
return Pred
```