

TS-correlation

Paolo Bosetti

2/7/2022

Correlazione tra due serie temporali

Questo esempio mostra come identificare il ritardo tra due serie temporali, della quali la seconda è un effetto ritardato della prima. È il caso, ad esempio, in cui si desidera stimare il ritardo con cui il segnale del livello di piena di un bacino si riflette sul livello dell'acqua di falda ad una certa distanza: il secondo segnale, visto come serie temporale, sarà simile al primo, ma ritardato e probabilmente filtrato nelle variazioni più brusche (cioè nelle componenti in frequenza più alta).

Costruiamo un esempio usando la funzione `arma.sim()` per creare una serie temporale `d1`. La funzione `arma.sim()` ritorna un oggetto `ts`, che per comodità convertiamo in `xts`:

```
set.seed(1)
lag <- 43
l1 <- 30
l2 <- -5
d1 <- l1 + arma.sim(list(ar=c(0.5, -0.2, 0.3, -0.1, 0.2),
                        ma=2,
                        order=c(5, 1, 1)),
                    600)
d1 <- xts(d1,
          order.by = seq(from=ISOdate(2022, 1, 1), by="min", along.with=d1))
```

La serie `d2` è creata addolcendo `d1` mediante un kernel smoothing con una banda da 10 minuti (600 secondi), ed è poi ritardata di 43 minuti rispetto a `d1`:

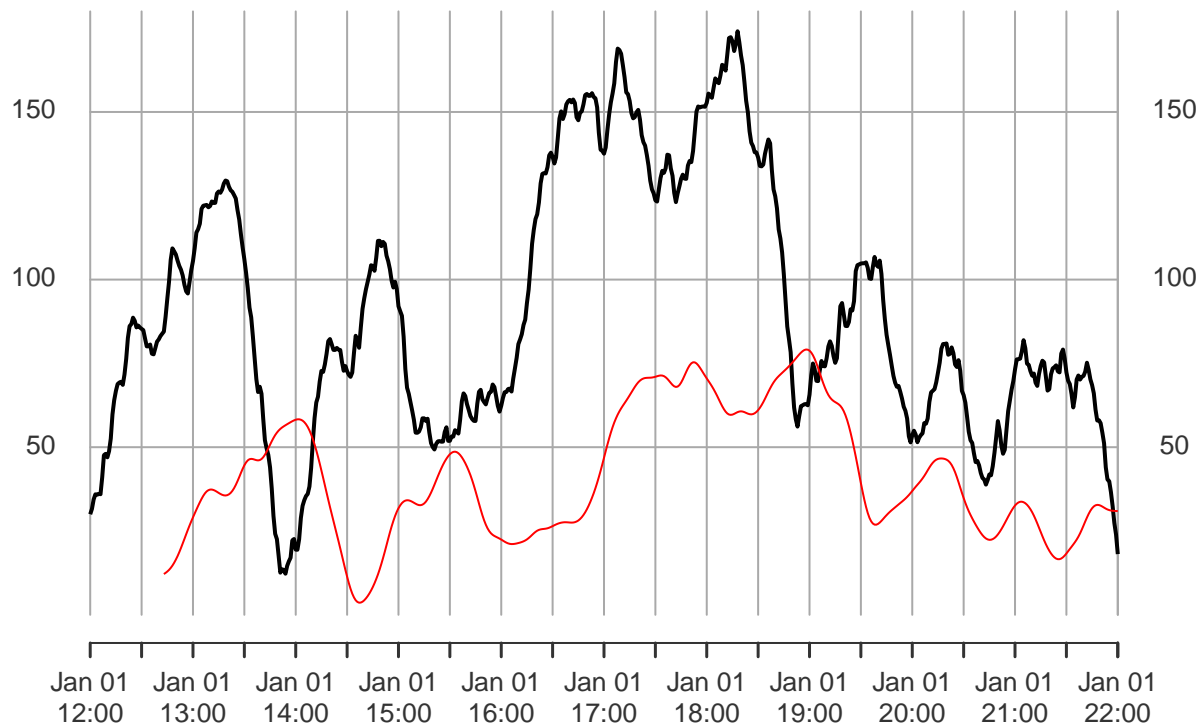
```
d2 <- l2 + ts(
  ksmooth(time(d1), d1, kernel="normal", bandwidth = 600)$y,
  start = start(d1)
) * 0.5
d2 <- xts(d2,
          order.by = seq(from=ISOdate(2022, 1, 1), by="min", along.with=d2))
d2 <- lag(d2, lag, na.pad=F)
```

Le due serie risultano essere:

```
invisible(plot(d1, typ="l", ylim=c(0, 180)))
lines(d2, col="red")
```

d1

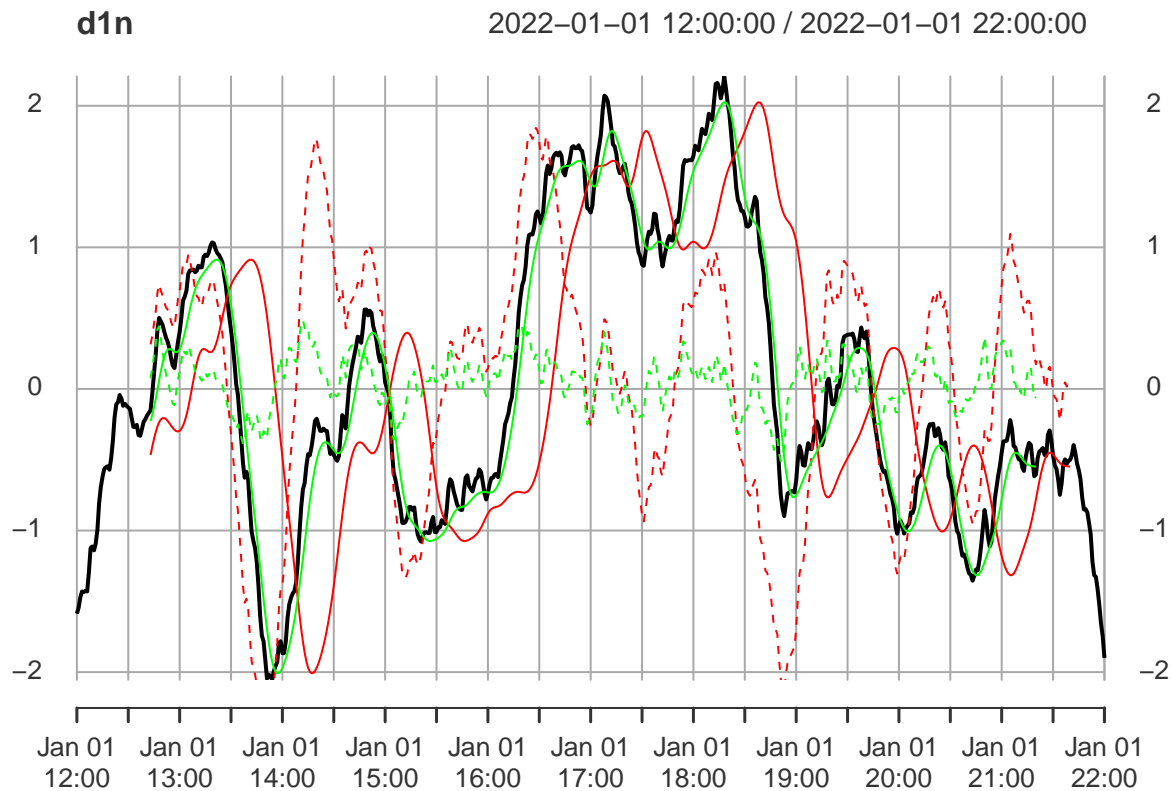
2022-01-01 12:00:00 / 2022-01-01 22:00:00



Ora fingiamo di non conoscere il ritardo (*lag*). Per identificarlo possiamo procedere come segue:

1. **normalizziamo** i due segnali: $x_n = (x - \bar{x})/\sigma_x$, che in questo modo diventano **confrontabili**;
2. valutiamo la differenza tra il primo segnale normalizzato e il secondo segnale, normalizzato e ritardato di un certo valore;

```
normalize <- function(x) (x - mean(x, na.rm = T))/sd(x, na.rm = T)
d1n <- normalize(d1)
d2n <- normalize(d2)
l <- 20
invisible(plot(d1n, typ="l"))
invisible(lines(lag(d2n, -l), col="red"))
invisible(lines(d1n - lag(d2n, -l), col="red", lty=2))
l <- 40
invisible(lines(lag(d2n, -l), col="green"))
lines(d1n - lag(d2n, -l), col="green", lty=2)
```



È evidente che la deviazione standard della differenza sarà tanto più piccola quanto più i due segnali saranno simili. Possiamo quindi esplorare un intero intervallo di ritardi e trovare il ritardo a cui corrisponde la minor deviazione standard:

```
lv <- 1:60
v <- sapply(lv, function(n) sd(d1n - lag(d2n, -n), na.rm = T))
plot(lv, v)
abline(v=lv[which.min(v)], col="gray")
text(lv[which.min(v)], 0.5, labels=paste("Lag:", lv[which.min(v)]), srt=90, pos=3)
```

