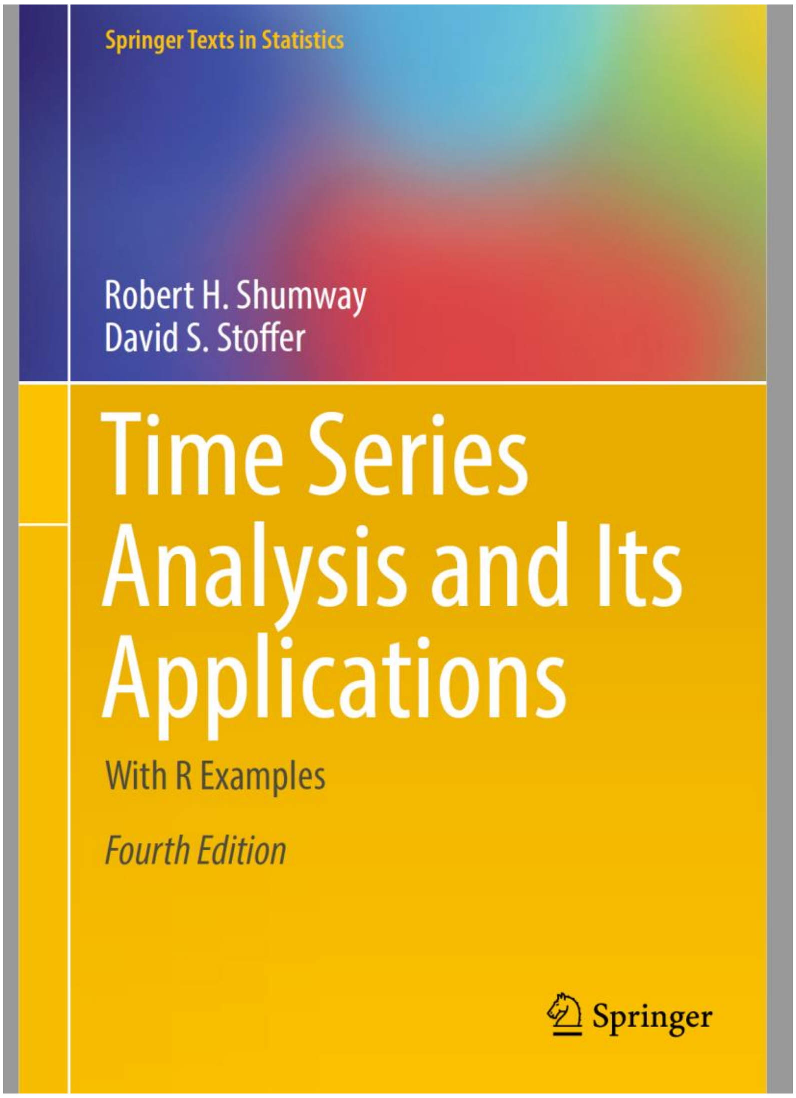
MAE 5870 - Aula 1



Série Temporal:

Uma série temporal é qualquer conjunto de observações ordenadas no tempo.

Example 1.1 Johnson & Johnson Quarterly Earnings

Figure 1.1 shows quarterly earnings per share for the U.S. company Johnson & Johnson, furnished by Professor Paul Griffin (personal communication) of the Graduate School of Management, University of California, Davis. There are 84 quarters (21 years) measured from the first quarter of 1960 to the last quarter of 1980. Modeling such series begins by observing the primary patterns in the time history. In this case, note the gradually increasing underlying trend and the rather regular variation superimposed on the trend that seems to repeat over quarters. Methods for analyzing data such as these are explored in Chapter 2 and Chapter 6. To plot the data using the R statistical package, type the following:[1.1]

```
library(astsa)     # SEE THE FOOTNOTE
plot(jj, type="o", ylab="Quarterly Earnings per Share")
```

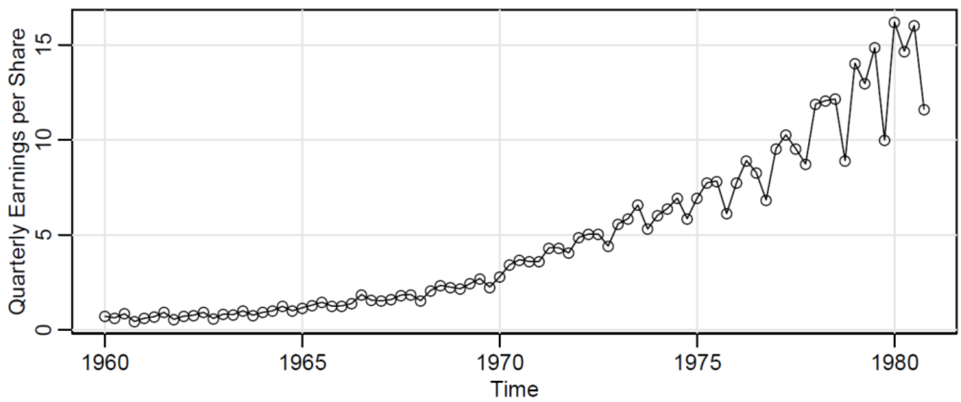


Fig. 1.1. Johnson & Johnson quarterly earnings per share, 84 quarters, 1960-I to 1980-IV.

Example 1.2 Global Warming

Consider the global temperature series record shown in Figure 1.2. The data are the global mean land–ocean temperature index from 1880 to 2015, with the base period 1951-1980. In particular, the data are deviations, measured in degrees centigrade, from the 1951-1980 average, and are an update of Hansen et al. (2006). We note an apparent upward trend in the series during the latter part of the twentieth century that has been used as an argument for the global warming hypothesis. Note also the leveling off at about 1935 and then another rather sharp upward trend at about 1970. The question of interest for global warming proponents and opponents is whether the overall trend is natural or whether it is caused by some human-induced interface. Problem 2.8 examines 634 years of glacial sediment data that might be taken as a long-term temperature proxy. Such percentage changes in temperature do not seem to be unusual over a time period of 100 years. Again, the question of trend is of more interest than particular periodicities. The R code for this example is similar to the code in Example 1.1:

```
plot(globtemp, type="o", ylab="Global Temperature Deviations")
```

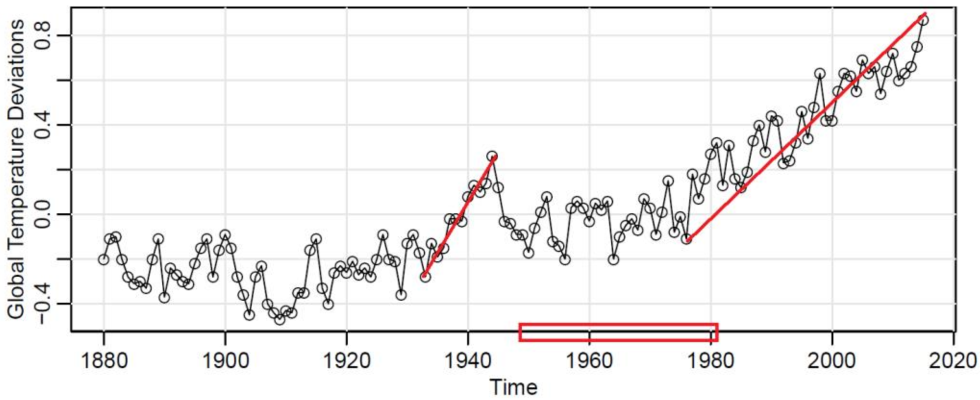


Fig. 1.2. Yearly average global temperature deviations (1880–2015) in degrees centigrade.

Example 1.3 Speech Data

Figure 1.3 shows a small .1 second (1000 point) sample of recorded speech for the phrase $aaa \cdots hhh$, and we note the repetitive nature of the signal and the rather regular periodicities. One current problem of great interest is computer recognition of speech, which would require converting this particular signal into the recorded phrase $aaa \cdots hhh$. Spectral analysis can be used in this context to produce a signature of this phrase that can be compared with signatures of various library syllables to look for a match. One can immediately notice the rather regular repetition of small wavelets. The separation between the packets is known as the *pitch period* and represents the response of the vocal tract filter to a periodic sequence of pulses stimulated by the opening and closing of the glottis. In R, you can reproduce Figure 1.3 using `plot(speech)`.

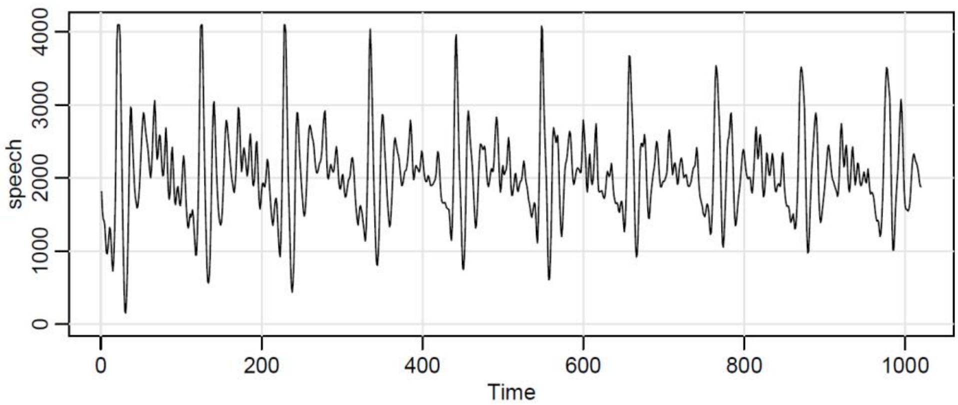


Fig. 1.3. Speech recording of the syllable $aaa \cdots hhh$ sampled at 10,000 points per second with $n = 1020$ points.

Example 1.4 Dow Jones Industrial Average

As an example of financial time series data, Figure 1.4 shows the $\boxed{\text{daily } \textit{returns}}$ (or percent change) of the Dow Jones Industrial Average (DJIA) from April 20, 2006 to April 20, 2016. It is easy to spot the financial crisis of 2008 in the figure. The data shown in Figure 1.4 are typical of return data. The mean of the series appears to be stable with an average return of approximately zero, however, highly volatile (variable) periods tend to be clustered together. A problem in the analysis of these type of financial data is to forecast the volatility of future returns. Models such as *ARCH* and *GARCH* models (Engle, 1982; Bollerslev, 1986) and *stochastic volatility* models (Harvey, Ruiz and Shephard, 1994) have been developed to handle these problems. We will discuss these models and the analysis of financial data in Chapter 5 and Chapter 6. The data were obtained using the Technical Trading Rules (TTR) package to download the data from Yahoo™ and then plot it. We then used the fact that if x_t is the actual value of the DJIA and $r_t = (x_t - x_{t-1})/x_{t-1}$ is the return, then $1 + r_t = x_t/x_{t-1}$ and $\log(1 + r_t) = \log(x_t/x_{t-1}) = \log(x_t) - \log(x_{t-1}) \approx r_t$.[1.2] The data set is also available in astsa, but xts must be loaded.

```
# library(TTR)
# djia  = getYahooData("^DJI", start=20060420, end=20160420, freq="daily")
library(xts)
djiar = diff(log(djia$Close))[-1] # approximate returns
plot(djiar, main="DJIA Returns", type="n")
lines(djiar)
```

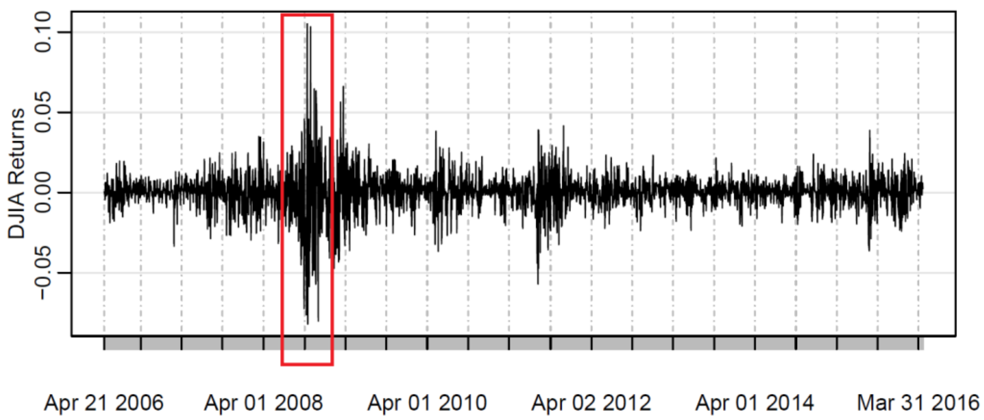


Fig. 1.4. *The daily returns of the Dow Jones Industrial Average (DJIA) from April 20, 2006 to April 20, 2016.*

Example 1.5 El Niño and Fish Population

We may also be interested in analyzing several time series at once. Figure 1.5 shows monthly values of an environmental series called the *Southern Oscillation Index* (SOI) and associated Recruitment (number of new fish) furnished byDr. Roy Mendelssohn of the Pacific Environmental Fisheries Group (personal communication). Both series are for a period of 453 months ranging over the years 1950–1987. The SOI measures changes in air pressure, related to sea surface temperatures in the central Pacific Ocean. The central Pacific warms every three to seven years due to the El Niño effect, which has been blamed for various global extreme weather events. Both series in Figure 1.5 exhibit repetitive behavior, with regularly repeating *cycles* that are easily visible. This periodic behavior is of interest because underlying processes of interest may be regular and the rate or *frequency* of oscillation characterizing the behavior of the underlying series would help to identify them. The series show two basic oscillations types, an obvious annual cycle (hot in the summer, cold in the winter), and a slower frequency that seems to repeat about every 4 years. The study of the kinds of cycles and their strengths is the subject of Chapter 4. The two series are also related; it is easy to imagine the fish population is dependent on the ocean temperature. This possibility suggests trying some version of regression analysis as a procedure for relating the two series. *Transfer function modeling*, as considered in Chapter 5, can also be applied in this case. The following R code will reproduce Figure 1.5:

```
par(mfrow = c(2,1))  # set up the graphics
plot(soi, ylab="", xlab="", main="Southern Oscillation Index")
plot(rec, ylab="", xlab="", main="Recruitment")
```

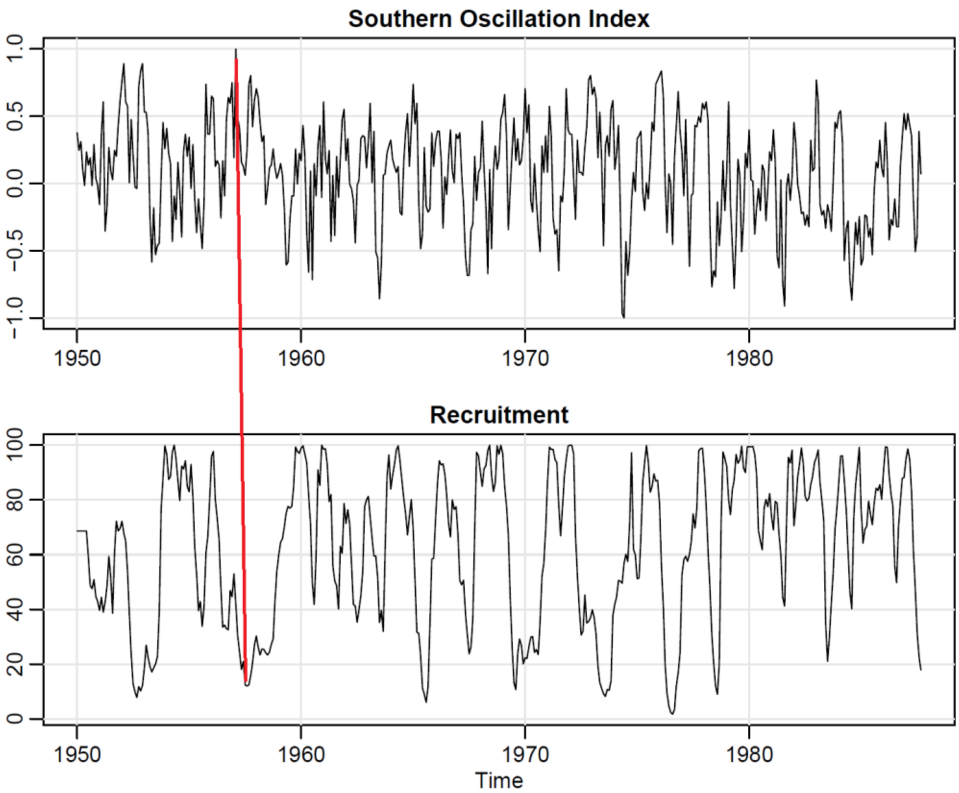


Fig. 1.5. Monthly SOI and Recruitment (estimated new fish), 1950-1987.

Example 1.6 fMRI Imaging

A fundamental problem in classical statistics occurs when we are given a collection of independent series or vectors of series, generated under varying experimental conditions or treatment configurations. Such a set of series is shown in Figure 1.6, where we observe data collected from various locations in the brain via functional magnetic resonance imaging (fMRI). In this example, five subjects were given periodic brushing on the hand. The stimulus was applied for 32 seconds and then stopped for 32 seconds; thus, the signal period is 64 seconds. The sampling rate was one observation every 2 seconds for 256 seconds ($n = 128$). For this example, we averaged the results over subjects (these were evoked responses, and all subjects were in phase). The series shown in Figure 1.6 are consecutive measures of blood oxygenation-level dependent (BOLD) signal intensity, which measures areas of activation in the brain. Notice that the periodicities appear strongly in the motor cortex series and less strongly in the thalamus and cerebellum. The fact that one has serie from different areas of the brain suggests testing whether the areas are respondin differently to the brush stimulus. Analysis of variance techniques accomplish this i classical statistics, and we show in Chapter 7 how these classical techniques exten to the time series case, leading to a spectral analysis of variance. The following I commands can be used to plot the data:

```
par(mfrow=c(2,1))
ts.plot(fmri1[,2:5], col=1:4, ylab="BOLD", main="Cortex")
ts.plot(fmri1[,6:9], col=1:4, ylab="BOLD", main="Thalamus & Cerebellum")
```

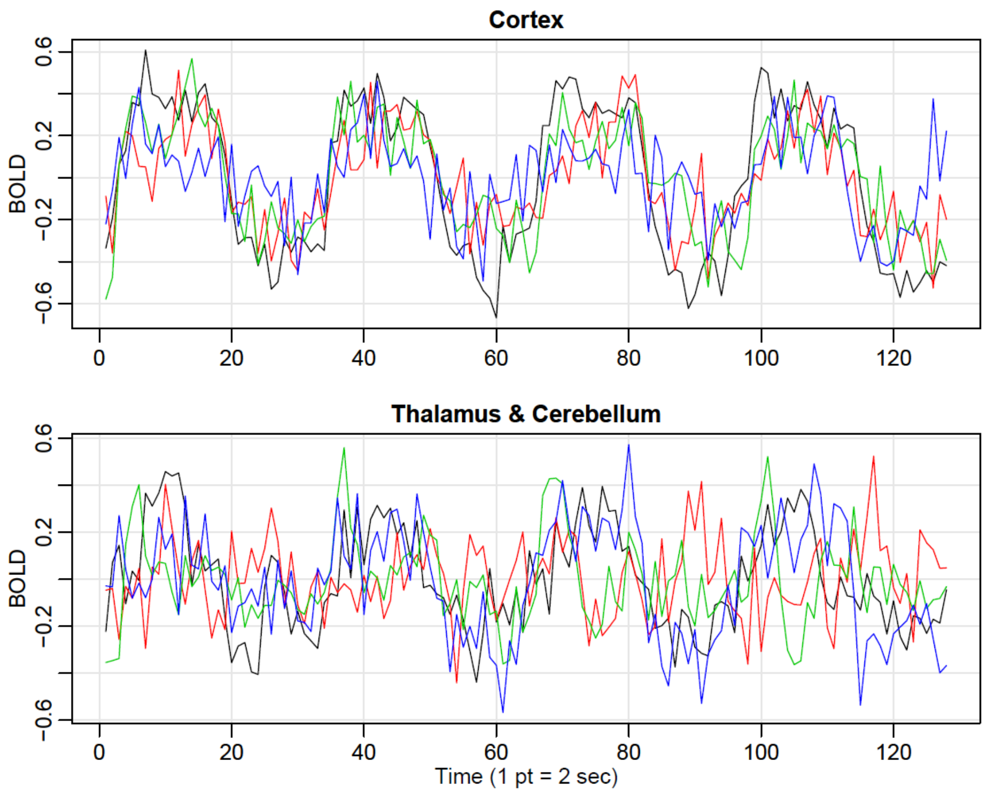


Fig. 1.6. *fMRI data from various locations in the cortex, thalamus, and cerebellum;* $n = 128$ *points, one observation taken every 2 seconds.*

Example 1.7 Earthquakes and Explosions

As a final example, the series in Figure 1.7 represent two phases or arrivals along the surface, denoted by P ($t = 1, \ldots, 1024$) and S ($t = 1025, \ldots, 2048$), at a seismic recording station. The recording instruments in Scandinavia are observing earthquakes and mining explosions with one of each shown in Figure 1.7. The general problem of interest is in distinguishing or discriminating between waveforms generated by earthquakes and those generated by explosions. Features that may be important are the rough amplitude ratios of the first phase P to the second phase S, which tend to be smaller for earthquakes than for explosions. In the case of the two events in Figure 1.7, the ratio of maximum amplitudes appears to be somewhat less than .5 for the earthquake and about 1 for the explosion. Otherwise, note a subtle difference exists in the periodic nature of the S phase for the earthquake. We can again think about spectral analysis of variance for testing the equality of the periodic components of earthquakes and explosions. We would also like to be able to classify future P and S components from events of unknown origin, leading to the *time series discriminant analysis* developed in Chapter 7.

To plot the data as in this example, use the following commands in R:

```
par(mfrow=c(2,1))
plot(EQ5, main="Earthquake")
plot(EXP6, main="Explosion")
```

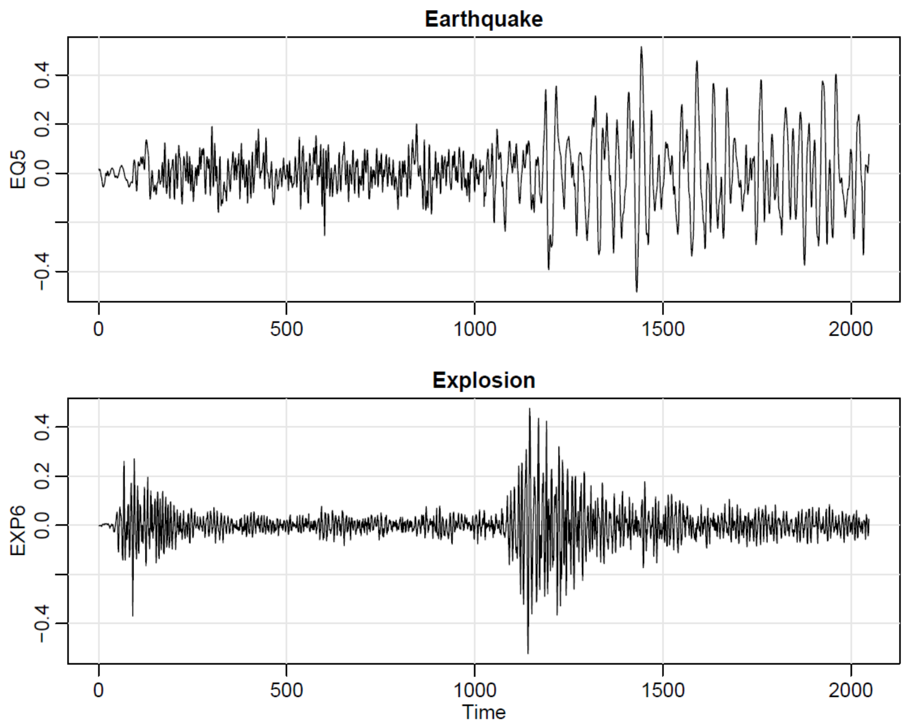


Fig. 1.7. Arrival phases from an earthquake (top) and explosion (bottom) at 40 points per second.

Objetivo da análise de séries temporais

- Investigar o mecanismo gerador da série temporal;
- Fazer previsões de valores futuros da série;
- Descrever apenas o comportamento da série: existência de tendência, ciclos e variações sazonais;
- Procurar periodicidades relevantes nos dados.

Ferramentas

- Descrever o comportamento da série: gráficos e testes para avaliar tendências, ciclos, variações sazonais;
- Inferências estatísticas;
- Modelagem do fenômeno estudado;
- Previsões.

Tipos de Séries Temporais

. **Discreta**: $X(t)$, $t=1,2, \ldots, n$
- Valores semanais do número de casos de Aids em São Paulo;
- Taxa de mortalidade (mensais, anuais);
- Gastos com a saúde (mensais, anuais).

. **Contínua**: $X(t)$,
- Valores do eletrocardiograma;
- Medições de temperatura e umidade.