

Blog

November 1st 2021



Weather forecast by air pressure?

Idea

A small anecdote beforehand: While commissioning the eHive in Salzburg (see previous blog post) and the subsequent function test of all sensors, a BeeBIT member based in Würzburg (approx. 180 m above sea level) immediately noticed the suspiciously low air pressure of only about 970 mbar. Values around 1000 mbar are typical for Würzburg. The measurement error of the barometer installed in the weather station is 1 mbar (see blog entry of 20th September 2019). After a brief moment of shock at the potentially defective sensor, however, relief quickly set in: of course, Salzburg's topography (approx. 425 m above sea level) is responsible for the comparatively low air pressure. According to the barometric formula, at an air pressure of 1013.25 mbar at sea level and a constant temperature of 15 °C, pressures of 992 mbar (Würzburg) and 963 (Salzburg) are to be expected. In clear weather, the measured values are therefore perfectly normal. In fact, the barometric altitude formula could be verified at all eHive sites by averaging the air pressure over a representative period and comparing it to the value expected by altitude. The exponential dependence between pressure and altitude could thus be guessed from a graphical plot of pressure versus altitude even without knowing the correct formula. However, such a procedure is not the subject of this blog post.

In the winter months, the bee colony has usually completely ceased flight and brood activities. The colony survives the cold season in the form of a so-called winter cluster. Although periodic temperature rises can often still be observed in the hive, which serve to liquefy and consume the honey reserves, otherwise there is comparatively little to be gained from the data. Of course, this winter dormancy only applies to the bee colony; the weather station mounted on the eHive continues to send interesting data that can be evaluated and interpreted. In the following paragraphs, an attempt will be made to establish a relationship between air pressure and solar radiation. The starting point is the weather symbols (sunshine, clouds, rain) often found on simple barometers, which reveal the purpose of these barometers: They are used to predict the weather. Following a simple figurative idea, low air pressure attracts clouds from surrounding regions, while high pressure displaces the clouds and thus produces »nice« weather. Can we confirm this relationship using eHive weather data? If so, how reliable is this method?

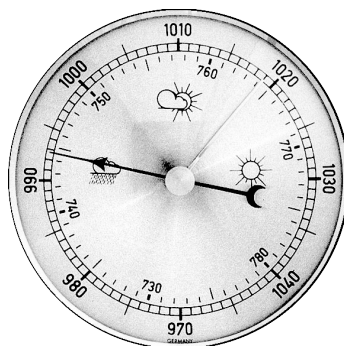


Fig.1: Barometer faceplate with symbols for weather forecasting.

The following investigation is the third contribution on the topic of data analysis after the articles

of July 20th 2019 and September 22nd 2019. Once again, the Python3 programming language is used, but the calculations and graphical representations performed can also be implemented in other tools such as spreadsheet programs. The Python script and raw data downloaded by means of diagram display are linked at the end of this article. In the script, you can follow all the calculations performed using the program instructions. The mathematical syntax follows the specifications of the NumPy library.

Data set 1: Autumn in Salzburg

Raw data over time

Following the above anecdote, we first evaluate data from the eHive in Salzburg shortly after its commissioning. In the period from October 14th to October 30th of this year, the data for air pressure and solar radiation were selected in the diagram display and downloaded by clicking on the download button. Data not needed at the beginning and end of the time range were deleted by hand, leaving 17 full days each from 00:00 to 23:59. All time data refer to the local summer time (UTC+2). Accordingly, the time range from 10/13 22:00 to 10/30 21:59 (UTC+0) is found in the data set. The raw data with daily mean values are plotted in Fig. 2.

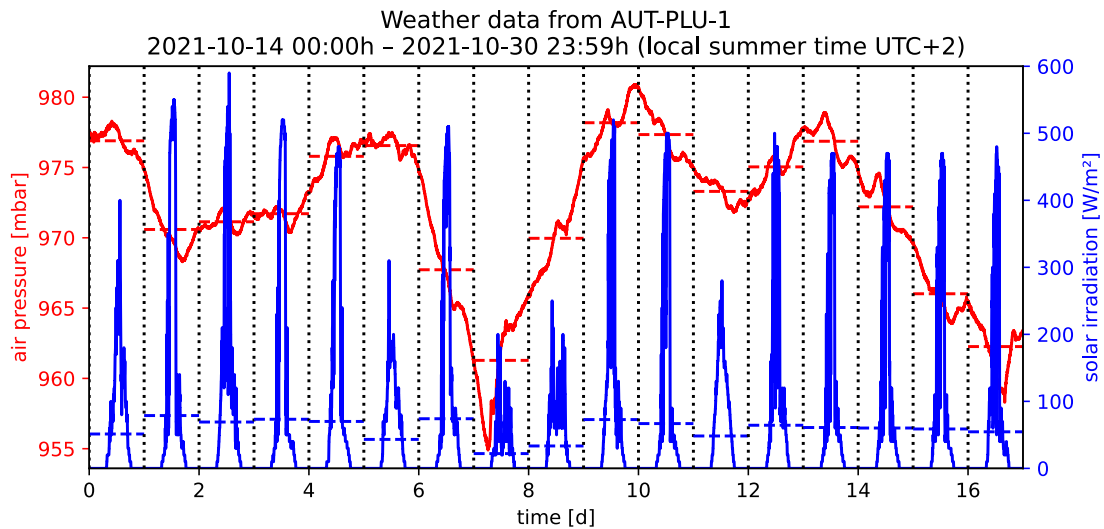


Fig.2: Air pressure and solar irradiance of the eHive AUT-PLU-1 in the time range from October 14th 2021 to October 30th 2021 (inclusive). Dotted vertical lines mark midnight in local time (UTC+2). Daily averages for pressure and solar radiation are plotted as dashed horizontal lines.

In principle, we can already check our working hypothesis with the help of Fig. 2. Striking, for example, is the strong drop in air pressure beginning on day 6 of the measurement period. While on this day the solar radiation still reaches values of up to 500 W/m^2 (daily mean approx. 70 W/m^2), the maximum value on day 7 drops to only about 200 W/m^2 (daily mean 20 W/m^2). A dip in the observed solar radiation indicates cloudy or even rainy weather. The same phenomenon under reversed conditions can be seen from day 7 to 9: While air pressure increases approximately linearly from 955 mbar to over 980 mbar , a slight increase in daily mean solar radiation can be observed from day 7 to 8 and a significant increase from day 8 to 9. But is the correlation between air pressure and solar radiation really that simple?

Recognizing correlations

From day 14 on, the pressure drops almost linearly from 980 mbar to about 960 mbar. Nevertheless, the solar radiation remains almost constant. It can be assumed that the effect observed on days 6 to 9 was only a coincidence. To be able to quantify the degree of correlation between the plotted measured variables, we need a new plot. Building on the blog post from 09/22/2019, the data points of both measured values are plotted against each other, see Fig. 3.

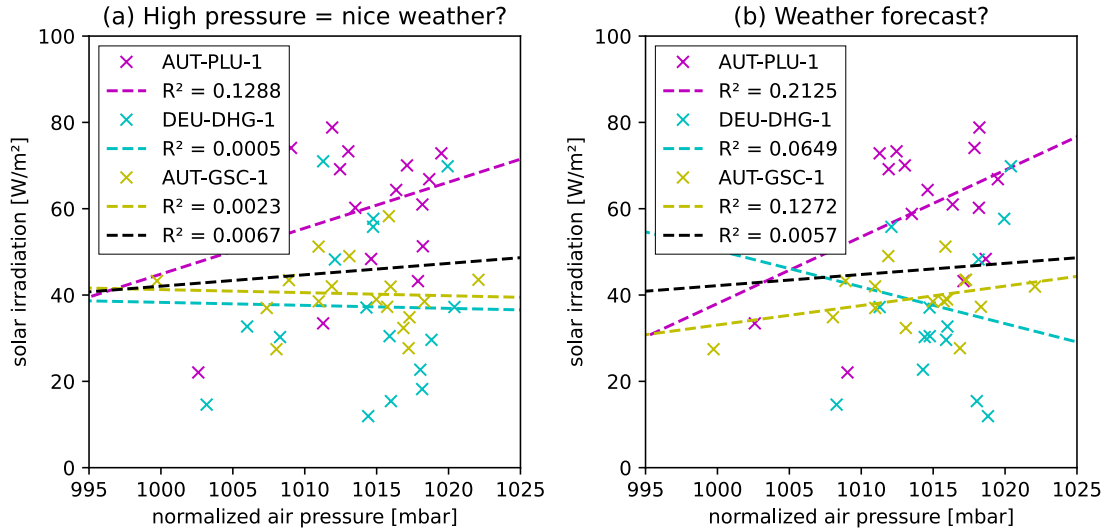


Fig. 3: Correlation between air pressure and solar radiation of the eHives AUT-PLU-1, DEU-DHG-1 and AUT-GSC-1 in the time range from 10/14/2021 to 10/30/2021 with (a) plotting of the daily mean values of the same day or with (b) autraging of the solar radiation against the air pressure of the previous day. A linear regression was calculated using the method of least squares and is drawn dashed. In the legend the value of the coefficient of determination R^2 of the linear fit can be found.

In Fig. 3, for better comparability with other locations, all air pressure values were normalized so that the mean value over the entire time range corresponds to the value of 1013.25 mbar expected at sea level. Although this normalization could also be performed using the barometric altitude formula, since the air pressure curves of geographically close locations (all Central Europe north of the Alps) almost duplicate each other except for the altitude offset, this method was chosen, which does not require any further data such as the exact altitude of the eHive and the outdoor temperature.

First, the daily mean values of air pressure and solar radiation of the same day were plotted as crosses in Fig. 3a. A perfect linear correlation would exist if all measured values lie on a straight line. Under the assumption "high air pressure means nice weather", a positive slope is to be expected for the straight line. In fact, the shape of the registered measuring points deviates strongly from a straight line. With a little good will, a concentration around the diagonal line rising from the lower left to the upper right can be detected in the point cloud, which corresponds to the hypothesis. But there can be no question of a clear linear correlation. With imagination, almost any kind of curve fitting could be done for the point cloud, c.f. this relevant XKCD comic.

Linear regression and coefficient of determination

Nevertheless, to obtain a quantitative measure of the quality of the correlation, a compensation line was calculated using the least squares method and also plotted in Fig. 3a. In fact, a positive slope is obtained. From the so-called coefficient of determination R^2 of the linear fit, the degree

of functional correlation between the examined measured values can be estimated. The coefficient of determination takes values between 0 and 1, where 0 corresponds to a completely uncorrelated distribution and 1 to a perfect linear relationship. For the regression line between air pressure and solar radiation a value of $R^2=12.9\%$ is obtained. (For limitations and criticism of the coefficient of determination for estimating and finding a correlation, we refer to the corresponding Wikipedia entry. At this moment, we are content to point out that a high value of the coefficient of determination can also result from a spurious correlation and that there may be other variables that have not yet been considered. A statement about the statistical significance of the results has also not been made as of now).

The low value of the coefficient of determination found in Fig. 3a is consistent with the observation that the measured values are strongly scattered in the diagram and it is difficult to recognize a trend. Of course, a plot of the measured values of the same day as in Fig. 3a does not correspond to the original hypothesis, according to which the weather can be predicted by means of barometers. For this purpose, e.g., the solar radiation of the following day must be compared with the current air pressure, see Fig. 3b. A similar picture as before results, but a slightly better concentration of the measuring points around the main diagonal can be observed. Accordingly, the calculated value of the coefficient of determination $R^2=21.3\%$ is higher. From the considered data set we can conclude within the framework of low statistical significance that the air pressure is apparently more suitable for the prediction of the solar radiation of the following day than of the current day (in agreement with the initial hypothesis and the simplified picture idea of sucked-in/displaced clouds). But how reliable is this conclusion really? Is it possibly a coincidence?

To answer this question, weather data from two other eHives in the same time range were first analyzed and can also be seen in Fig. 3. For both additional eHives we proceeded analogously to the previous data set, including normalization of the air pressure averages. We now observe a completely different picture: Although the R^2 values in Fig. 3b are still higher than in 3a, the absolute values are significantly lower. There are even equalization lines with negative slopes. In addition, a compensation line was drawn through the data points of all eHives (dashed black), which with $R^2<1\%$ does not suggest a functional relationship. So are barometers with painted weather symbols bunk? This needs a more precise analysis.

First of all, it can be stated that the selected time range at the end of October is suboptimal. Towards the end of the year, the solar radiation decreases naturally. This effect was not compensated for and, with otherwise constant parameters, ensures a scattering of the measured values in the vertical direction. The size of the data set is also very small with 17 days. Thus, a statistically reliable statement can hardly be achieved. In addition, weather forecasting is known to be very complicated and air pressure is only one of many relevant parameters. It would be conceivable to improve the model by, for example, considering air humidity in addition to pressure. However, in order not to overcomplicate the presented analysis, we continue to restrict ourselves to air pressure as the only input parameter and instead extend the time domain. By selecting a data set around the summer solstice (June 22nd), we minimize the effect of seasonal changes.

Data set 2: Summer solstice in Würzburg

We consider the period from June 29th to July 16th 2021 of the eHive DEU-DHG-1 in Würzburg. Analogous to the previous procedure, the measured values were plotted against each other and a linear regression was calculated, see Fig. 4. First of all, the significantly higher absolute values of the solar radiation due to the solar maximum at the beginning of the summer are striking. Again, the higher R^2 value is obtained in Fig. 4b (i.e. when using the atmospheric pressure to predict the solar irradiance of the following day), even though the absolute value is still comparatively low with $R^2=16.6\%$. However, due to the larger data set in comparison to Salzburg a coincidental correlation can rather be excluded.

Again, in Fig. 4 data of two further eHives from the same time range are displayed. Especially in Fig. 4b it can be seen that the correlation between air pressure and solar radiation is now no longer a coincidence, but can be observed in all data sets. For the eHive DEU-MNG-1 (location Mönchengladbach) a comparatively high coefficient of determination $R^2=27.8\%$ is obtained. The

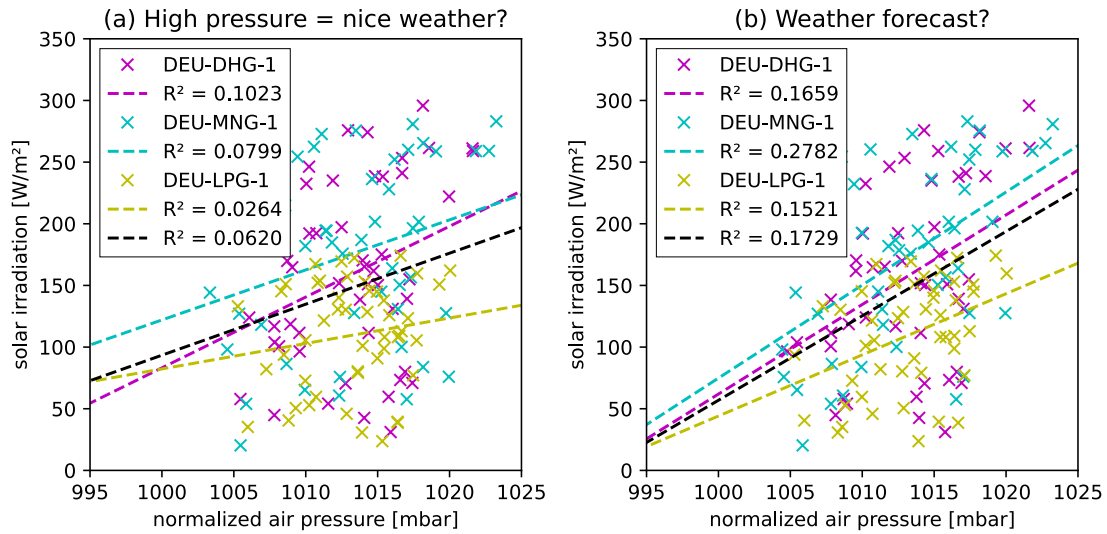


Fig.4: Correlation between air pressure and solar radiation of the eHives DEU-DHG-1, DEU-MNG-1 and DEU-LPG-1 in the time range from 05/29/2021 to 07/16/2021 by (a) plotting the daily mean values of the same day or by (b) plotting the solar radiation against the air pressure of the previous day. A regression line was calculated with the method of least squares and is drawn dashed. In the legend the value of the coefficient of determination R^2 of the linear fit can be found.

consistently lower values of solar radiation in Munich (DEU-LPG-1) are striking. It is possible that the weather station is temporarily shaded by a nearby building or tree.

Conclusion and outlook

We conclude this blog post with the realization that the correlation between air pressure and »nice« weather is by no means as linear as some simple barometers suggest. Nevertheless, a weak correlation was found after the methodology of the evaluation was reconsidered. Here, in particular, the selection of a sufficiently large time range at summer solstice have imposed themselves.

The data analysis carried out can, of course, only provide an insight into the subject matter. A multitude of further hypotheses can easily be formulated. For example, what is the effect of the air humidity already mentioned above? Is it perhaps useful to plot the change in air pressure from day to day (first time derivative) instead of the absolute value? Can a more precise prediction be made when comparing different locations? Such questions and others can be addressed in an analogous way. The size of the data set and the complexity of the methodology can be scaled almost arbitrarily.

The script used and used raw data can be found in the appendix of this blog post. If you have any questions, suggestions or criticism about this post or the project in general, please feel free to contact us at any time.

Supplementary materials

- raw data (data_weather.zip)
- Python script (weather_forecast.py)