

Export und Auswertung von Rohdaten

Die Diagrammanzeige der BeeBIT-Datenbank unterstützt den Export von Rohdaten. Hierdurch wird eine tiefere Analyse der Daten möglich. Im Folgenden wird beispielhaft eine Untersuchung von Temperatur- und Gewichtsdaten des Bienenstocks AUT-GSC-1 durchgeführt. Zur besseren Nachvollziehbarkeit sind Datenexport und die anschließenden Analyseansätze in Schritte gegliedert. Die Datenverarbeitung erfolgt mit Python. Das vollständige Python-Skript sowie eine erläuternde PDF sind am Ende des Textes verlinkt. Ziel dieses Beitrags ist es, Inspiration und einige Ansätze für das eigenständige Arbeiten mit den Datensätzen zu liefern. Denn so viel vorweg: Häufig lohnt sich ein zweiter Blick.

1) Export der Daten aus der Diagrammanzeige

In der Diagrammanzeige unter bee-bit.de/diagram können die zu untersuchenden Datensätze komfortabel ausgewählt und betrachtet werden. Wir werden uns beispielhaft die Stocktemperatur (6 Sensoren an verschiedenen Positionen innerhalb des Bienenstocks), die Gewichtsdaten sowie die Außentemperatur des eHives AUT-GSC-1 vom 10. Juni bis einschließlich 12. Juni 2019 näher ansehen. Hierbei ist zu beachten, dass alle Zeiten in UTC+0 angegeben werden, d.h. die tatsächliche Ortszeit (MESZ = UTC+2) ist um 2h verschoben. Wir möchten also eigentlich den Zeitraum vom 09.06. 22 Uhr bis 12.06. 22 Uhr exportieren, um in der lokalen Zeit den Drei-Tages-Zeitraum von Mitternacht bis Mitternacht zu erhalten.

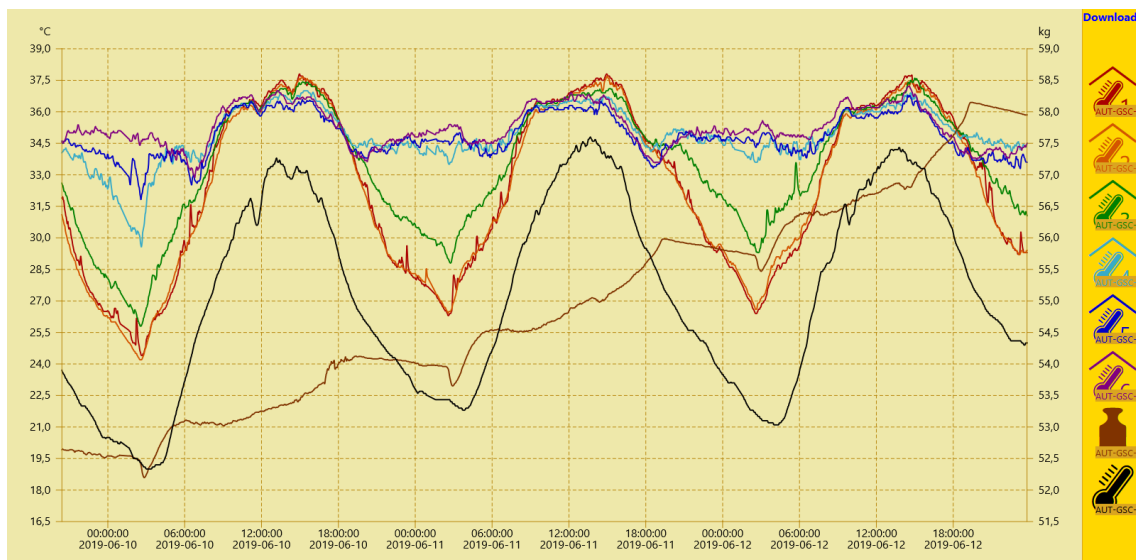


Abb.1: Stocktemperatur, Gewicht und Außentemperatur des eHives AUT-GSC-1 im Zeitraum 09.06.2019, 22 Uhr bis 12.06.2019, 22 Uhr (UTC+0) dargestellt über die Diagrammanzeige der BeeBIT-Website.

Durch Klick auf den Download-Button in der oberen rechten Ecke wird vom Server eine csv-Datei mit den ausgewählten Datensätzen generiert und zum Download bereitgestellt. Diese Datei kann von Texteditoren oder auch in Tabellenkalkulations-Programmen geöffnet werden. Da wir im

Folgenden jedoch auch rechenlastige Aufgaben zur Verarbeitung der Daten durchführen wollen, soll der Datensatz mit einem Skript in der high-level Programmiersprache Python eingelesen, ausgewertet und dargestellt werden.

Wir öffnen die Daten dennoch zunächst in einem Texteditor, um die Struktur des Datensatzes erkennen zu können. Der standardmäßige Name der heruntergeladenen Datei lautet data.csv. In der ersten Zeile sind die Namen der in der Datei abgelegten Rohdaten bezeichnet. In den weiteren Zeilen finden wir die Rohdaten, jeweils pro Zeitschritt durch einen Zeilenumbruch getrennt. Innerhalb einer Zeile wird das Semikolon als Trennzeichen genutzt. Hierdurch können wir den Datensatz gedanklich in Zeilen und Spalten gliedern, siehe Snippet 1. Die ersten beiden Spalten einer jeden aus der Diagrammanzeige heruntergeladenen csv-Datei beinhalten die Zeitinformation, einmal als Unix-Code in Dezimaldarstellung, einmal in menschenlesbarer Darstellung. In den folgenden Spalten befinden sich die in der Diagrammanzeige ausgewählten Sensordaten. Falls sich in der heruntergeladenen Datei noch Daten vor oder nach dem 72h Zeitbereich von 22 Uhr bis 22 Uhr (UTC+0) befinden, können diese von Hand herausgelöscht werden.

Snip.1: Die ersten und letzten drei Zeilen mit Rohdaten aus der heruntergeladenen Datei data.csv. Nicht dargestellt sind die Namen der Datensätze in Zeile 1.

```

2 | 1560117600;Sun, 09 Jun 2019 22:00:00 GMT;28,5;28;30;34;34,3;34,9;52,61;22
3 | 1560117660;Sun, 09 Jun 2019 22:01:00 GMT;28,5;28;30;33,9;34,2;34,9;52,61;22
4 | 1560117720;Sun, 09 Jun 2019 22:02:00 GMT;28,4;27,9;30;33,9;34,2;34,9;52,61;22

4205 | 1560376620;Wed, 12 Jun 2019 21:57:00 GMT;30;30,1;32,4;34,5;33,9;34;58,04;25,7
4206 | 1560376680;Wed, 12 Jun 2019 21:58:00 GMT;30;30,1;32,4;34,5;33,8;34;58,03;25,7
4207 | 1560376740;Wed, 12 Jun 2019 21:59:00 GMT;29,9;30,1;32,4;34,5;33,8;34;58,03;25,6

```

2) Darstellung der Rohdaten mit Python

Nachdem wir die Struktur des Datensatzes erfasst haben, können wir versuchen, die Daten durch den Computer einlesen zu lassen und sie zunächst einmal unverändert darzustellen. Im Folgenden verwenden wir hierfür die kostenlose Programmiersprache Python 3 mit den Bibliotheken NumPy und Matplotlib. Python ist plattformübergreifend und vergleichsweise einfach zu lernen. Die hier vorgetragene Analysemethode könnte z.B. in Form einer Projektarbeit auch von technikaffinen Schülern erlernt, angewendet und durch eigene Ansätze erweitert werden.

Zunächst wird die Datei eingelesen, die Datensätze in NumPy-Arrays gespeichert und die Sensordaten zeitaufgelöst mit Matplotlib geplottet, siehe Abb.2. Weil uns nur die relative Zeit interessiert, können wir die horizontale Zeitachse in Einheiten von Stunden gliedern, wobei 0h dem 09.06.2019 um 22 Uhr (UTC+0, Ortszeit in UTC+2: 10.06. 0 Uhr) entspricht. Da der Datensatz sowohl Temperatur als auch Gewichtsdaten enthält, werden zwei vertikale Achsen benötigt: eine in Einheiten von °C, die andere in Einheiten von kg. Eine Legende hilft, die farbcodiert geplotteten Datensätze den einzelnen Sensoren zuzuordnen. Zur besseren Orientierung wurden noch vertikale schwarze Linien im Abstand von 24h eingefügt, die den Tageswechsel um Mitternacht in Ortszeit markieren.

Der schwerste Schritt ist damit geschafft. Nachdem wir die Struktur des Datensatzes verstanden haben und die Daten korrekt darstellen konnten, gehen wir nun zur Analyse über. Vorher sei allerdings angemerkt, dass wir bereits durch einen genauen Blick auf die Temperaturverläufe der Stockinnensensoren eine Erkenntnis über die Position des Bienenvolkes innerhalb des eHives erhalten. Wie den geplotteten Verläufen zu entnehmen ist, bleibt die Temperatur der Sensoren 4 bis 6 annähernd konstant und schwankt im Tagesverlauf nur um ca. 3 °C um einen Mittelwert von ca. 35 °C. Diese Temperatur entspricht genau dem Wert, den das Bienenvolk zur Brut benötigt. Im Gegensatz dazu beträgt die Amplitude der Schwankung bei den Sensoren 1 bis 3 nahezu 10 °C. Dies lässt darauf schließen, dass sich das Volk hauptsächlich in der Nähe der Sensoren 4 bis 6 aufhält.

3) Untersuchung der Temperaturregulation

Wir möchten die beobachtete Temperaturregulation nun genauer untersuchen, als dies noch mit der Diagrammanzeige der BeeBIT-Website möglich war. Interessant ist der Unterschied zwischen Stockinnentemperatur und Außentemperatur abhängig von der Position innerhalb des eHives. Es werden beispielhaft die beiden Innensensoren 2 und 5 ausgewählt und die Innen-Außen-Differenz geplottet. Da wir eine allgemeine Aussage treffen möchten, die von einzelnen Ereignissen während der Datenerfassung möglichst unabhängig sein soll, mitteln wir über die drei Beobachtungstage, d.h. wir generieren aus jeweils drei Datenpunkten der drei Tage einen gemittelten Wert. Hierbei kommt uns die schnelle und intuitive Datenverarbeitung der NumPy-Bibliothek zu Hilfe. Der volle 72h Datensatz sowie der für eine Periode von 24h gemittelte Datensatz sind in Abb.3 geplottet, wobei für beide eine gemeinsame Temperatur-Achse verwendet wird.

Aus den dargestellten Innen-Außen-Differenzen wird deutlich, dass das Bienenvolk die Stocktemperatur reguliert. Wir haben bereits argumentiert, dass das Volk sich um Sensor 5 herum aufhält. Nachts wird das Stockinnere geheizt, damit die Temperatur nicht wie in der Nähe von Sensor 2 den Außentemperaturen folgt und es für die Brut im Stock zu kalt werden würde. Zwischen 5 Uhr und 7 Uhr erreicht die Innen-Außen-Differenz im 3-Tages-Mittel bei Sensor 5 ein Maximum von ca. 14 °C. Ganz anders sieht es in den warmen Nachmittags- und Abendstunden aus: Hier sinkt die Temperaturdifferenz bei Sensor 5 unter die Werte bei Sensor 2. Wenn wir davon ausgehen, dass der Temperaturverlauf bei Sensor 2 den Außentemperaturen (bis auf einen konstanten Offset) nahezu ungestört folgt, so können wir schließen, dass das Bienenvolk um Sensor 5 die umgebenden Waben aktiv kühlt. Tatsächlich ist dieses Verhalten den Imkern bekannt. Es wurde beobachtet, dass Arbeiterinnen Wasser in den Stock eintragen. Dieses kann verdunsten und der Stock bleibt kühl. Wir dürfen also vermuten, dass wir dieses interessante Phänomen in den untersuchten Daten beobachten konnten!

4) Untersuchung der Gewichtänderungsrate

Nach der Untersuchung der Temperaturregulation im eHive möchten wir abschließend noch den Gewichtsverlauf näher betrachten. Wie bereits in der Diagrammanzeige der BeeBIT-Website zu erkennen ist, steigt das Stockgewicht im Verlauf der drei Tage von ca. 52.5 kg auf ca. 58.0 kg. Das Absolutgewicht des Stocks beinhaltet jedoch das Gewicht sämtlicher Komponenten (Zarge, Sensoren, Technik, Biomasse). Interessanter ist eine Größe, die die relative Änderung des Gewichts dokumentiert: die Änderungsrate, d.h. die Zeitableitung des Gewichtsverlaufs. Diese Größe kann berechnet werden, indem wir die Differenz zweier benachbarter Datenpunkte bilden und durch die Dauer des Zeitschritts teilen. Plotten wir die Änderungsrate (Abb.4, oben) so stellen wir fest, dass das Messsignal sehr verrauscht ist. Zwar wird das Rauschen geringer, wenn erneut das Drei-Tages-Mittel berechnet wird, doch können wir mit diesen Daten noch keine Aussage über den Gewichtsverlauf treffen.

Wünschenswert wäre eine Bearbeitung, die das Rauschen aus dem Signal entfernen kann, ohne die Aussage der Messwerte allzu sehr abzuschwächen. Eine bekannte Methode zur Rauschunterdrückung in der Signalverarbeitung ist die Anwendung von Signalfiltern. Wir werden den sog. Gauß-Filter auf das Signal anwenden. Hierzu werden aus den Datenpunkten Interpolationspunkte berechnet. Die einzelnen Datenpunkte, die zur Berechnung eines Interpolationspunkts genutzt werden, werden jedoch unterschiedlich stark gewichtet. Ein Datenpunkt der zeitlich nahe dem Interpolationspunkt liegt, wird stärker gewichtet als ein Datenpunkt viele Minuten oder gar Stunden entfernt. Die Gewichtung entspricht hierbei einer Gaußschen Glockenkurve, die über dem jeweiligen Interpolationspunkt zentriert ist. Details zur Interpolation und dem verwendeten Algorithmus sind am Ende des Beitrags in Form einer erläuternden PDF-Datei verlinkt.

In den geglätteten Kurven (Abb.4, unten) können wir erkennen, dass insbesondere am Abend zwischen 16 Uhr und 20 Uhr das Gewicht des Stocks mit bis zu 0.3 kg/h stark ansteigt. Insgesamt ist eine positive Bilanz der Änderungsrate zu beobachten, d.h. im Tagesmittel ist die Änderungsrate größer null. Dies führt effektiv zu einer Zunahme des Stockgewichts, die wir bereits zu Beginn der Analyse feststellen konnten. Überraschend ist der starke Ausschlag am frühen Morgen: Nachdem

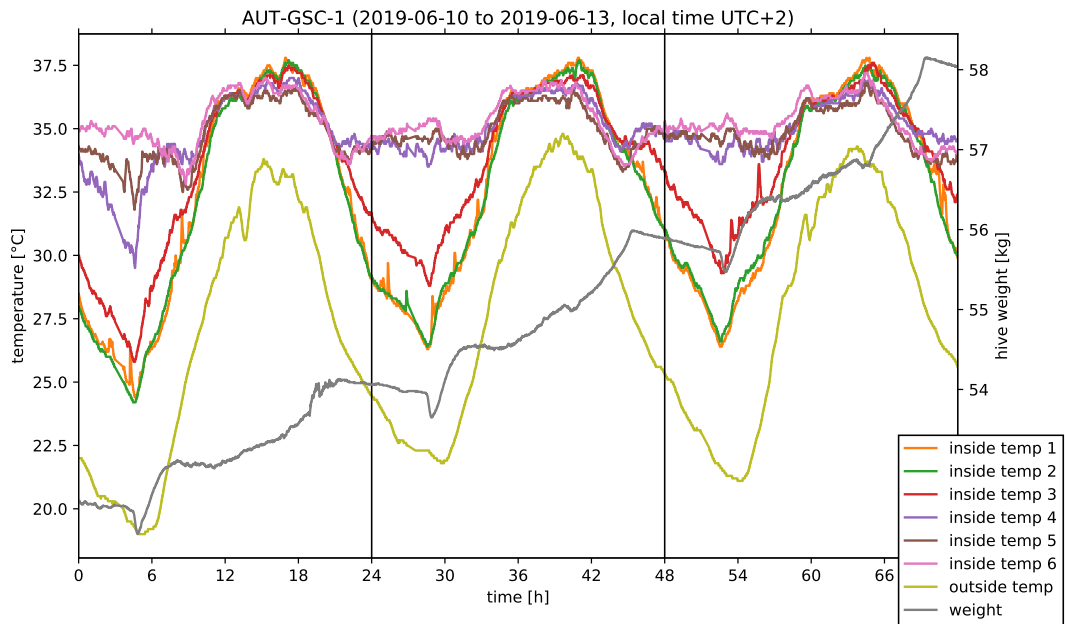


Abb.2: Darstellung des Datensatzes mit Python unter Verwendung der Bibliothek Matplotlib.

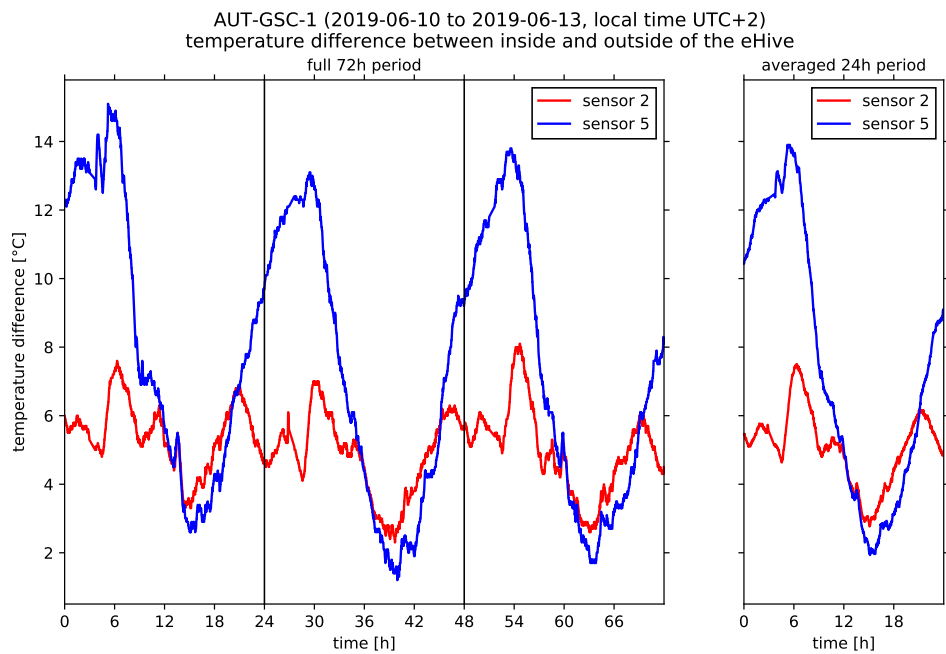


Abb.3: Temperatur-Differenz zwischen innerhalb und außerhalb des eHives für die beiden Innen-sensoren 2 und 5. Links die Rohdaten des dreitägigen Messzeitraums, rechts die gemittelten Tageswerte.

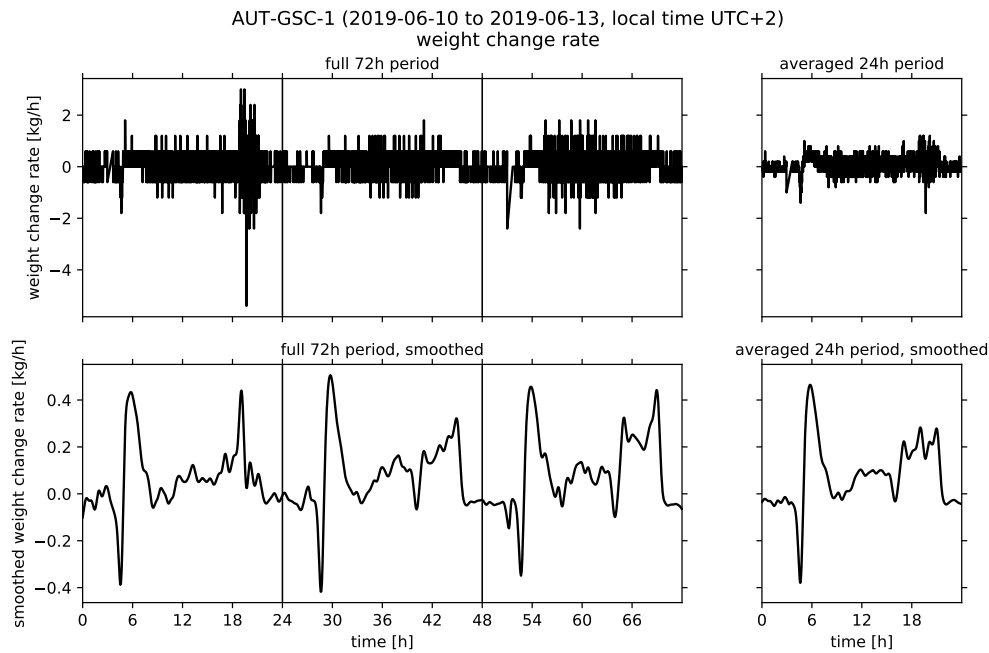


Abb.4: Gewichtänderungsrate des eHives. Links jeweils die Daten des dreitägigen Messzeitraums, rechts die gemittelten Tageswerte. Oben der ungeglättete Datensatz dargestellt mit gemeinsamer Raten-Achse, unten nach Anwendung des Gauß-Filters ($\sigma=0.25h$) mit gemeinsamer Raten-Achse aber anderer Skalierung als bei den ungefilterten Daten.

das Gewicht zuerst mit bis zu -0.4 kg/h stark abnimmt, steigt es kurz darauf erneut mir bis zu $+0.5$ kg/h an. Um 9 Uhr morgens ist die Änderungsrate dann wieder auf ca. 0.0 kg/h gesunken. Der Effekt kann an allen drei Tagen des Beobachtungszeitraums festgestellt werden und findet sich entsprechend auch im Drei-Tages-Mittel wieder. Meikle et al.¹ interpretierten das starke Absinken bei Tagesanbruch als Aufwachphase des Bienenvolkes. Während die Arbeiterinnen sich für den ersten Ausflug vorbereiten (und dabei Nahrung verbrauchen) und dann nach und nach den Stock verlassen, sinkt das Gewicht. Der darauffolgende kurzzeitige starke Anstieg ist in der Publikation von Meikle et al. jedoch so nicht zu finden. Vorstellbar wäre z.B., dass die Bienen Wasser in den Stock eintragen. Vielleicht hat die Gewichtsänderung aber auch keine biologische Ursache, sondern entsteht durch sich auf dem Stockäußeren niederschlagenden Morgentau. Ein schnelles Nachsehen in der Diagrammanzeige der BeeBIT-Website könnte diese zweite Vermutung bekräftigen: Genau zu der Zeit, zu der die Gewichtänderungsrate nach dem Aufwachprozess stark ansteigt, erreicht die Außenfeuchte an allen drei Tagen des Beobachtungszeitraums ein Maximum mit Werten von bis zu 80 %.

5) Zusammenfassung

Wir haben einen Datensatz aus der BeeBIT-Diagrammanzeige exportiert und mit einem Python-Skript verarbeitet und dargestellt. Insbesondere die Temperaturregulation im Stockinneren sowie der Gewichtsverlauf wurden näher untersucht. Es wurden numerisch die Differenz zweier Innentemperatur-Datensätze zur Außentemperatur, sowie die Zeitableitung des Gewichtsverlaufs gebildet. Auf letztere wurde ein Gauß-Filter angewendet, um das Rauschen des Messsignals zu unterdrücken. Die gezeigten Methoden sollen exemplarisch für denkbare Analyseansätze stehen. Solche Ansätze werden in der wissenschaftlichen Forschung genutzt, können jedoch auch von Schü-

¹ Meikle et al. (2018), PLoS ONE 13(5), *Using within-day hive weight changes to measure environmental effects on honey bee colonies*, <https://doi.org/10.1371/journal.pone.0197589> (open access)

lern und Laien entwickelt werden. Die eHives stellen hierbei interessante und anschaulich interpretierbare Datensätze zur Verfügung. Die durchgeführten Datenanalysen können in Komplexität und Schwierigkeitsgrad beliebig variiert werden. Auch die Arbeit mit nur einem kleinen Datensatz in einem leicht zu bedienenden Tabellenkalkulations-Programm kann bereits neue Erkenntnisse hervorbringen.² Tiefergehende numerische Analysen, wie sie z.B. in Projektarbeiten von älteren Schülern oder Studenten durchgeführt werden können, erlauben eine eingehende Beschäftigung mit den biologischen Aspekten der gesammelten Daten sowie den technischen Methodiken.

Wenn Sie, lieber Leser, Interesse bekommen haben, dann probieren Sie es gleich selbst einmal aus. Laden Sie sich einen Datensatz aus der Diagrammanzeige herunter, öffnen Sie ihn mit einem Programm Ihrer Wahl und schauen Sie genauer hin, als das noch auf der Website möglich war. Wie zu Beginn des Beitrags versprochen: Häufig lohnt sich ein zweiter Blick.

Wenn Sie Hilfe benötigen, Fragen zu diesem Beitrag haben oder sogar selbst schon einige Erkenntnisse anhand der Daten gewinnen konnten, dürfen Sie gerne Kontakt zu uns aufnehmen.

Ergänzende Materialien

- verwendete Rohdaten (data.csv)
- vollständiges Python-Script (export_and_analysis.py)
- Erläuterungen zur Methodik und deren Implementation (export_and_analysis_supportingmaterial.pdf, nur in englischer Sprache verfügbar)

² Ein für den Schulunterricht geeignetes Beispiel ist auf der BeeBIT-Website im Bereich der Unterrichtsmaterialien unter dem Titel »Datenauswertung mit einfachen Funktionen und Diagrammen« zu finden. Dort werden Daten zu Außentemperatur und Niederschlag in einer Excel-Arbeitsmappe aufbereitet und visualisiert.