

Simulations-Szenarien für den Wasserstand des Neusiedler Sees im Jahr 2020

Gerhard Svolba, Data Scientist
SAS, DACH Region



My Data Science Webinar Series:

[Youtube](#) | [Medium](#) | [LinkedIn](#)

<https://twitter.com/gsvolba>


<https://github.com/gerhard1050>

<https://www.linkedin.com/in/gerhardsvolba/>



Überblick

- Der fachliche Hintergrund: Wo? Wann? Warum?
- Datenaufbereitung für die Analyse
- Regressionsmodell für die Analyse der Zusammenhänge
- Concatenation Modell für eine einfache visuelle/interaktive Analyse



DATA SCIENCE CLASS
SIMULATIONS-SZENARIO FÜR DEN WASSERSTAND DES NEUSIEDLER SEES
23:44

Simulations-Szenarien für den Wasserstand des Neusiedler...

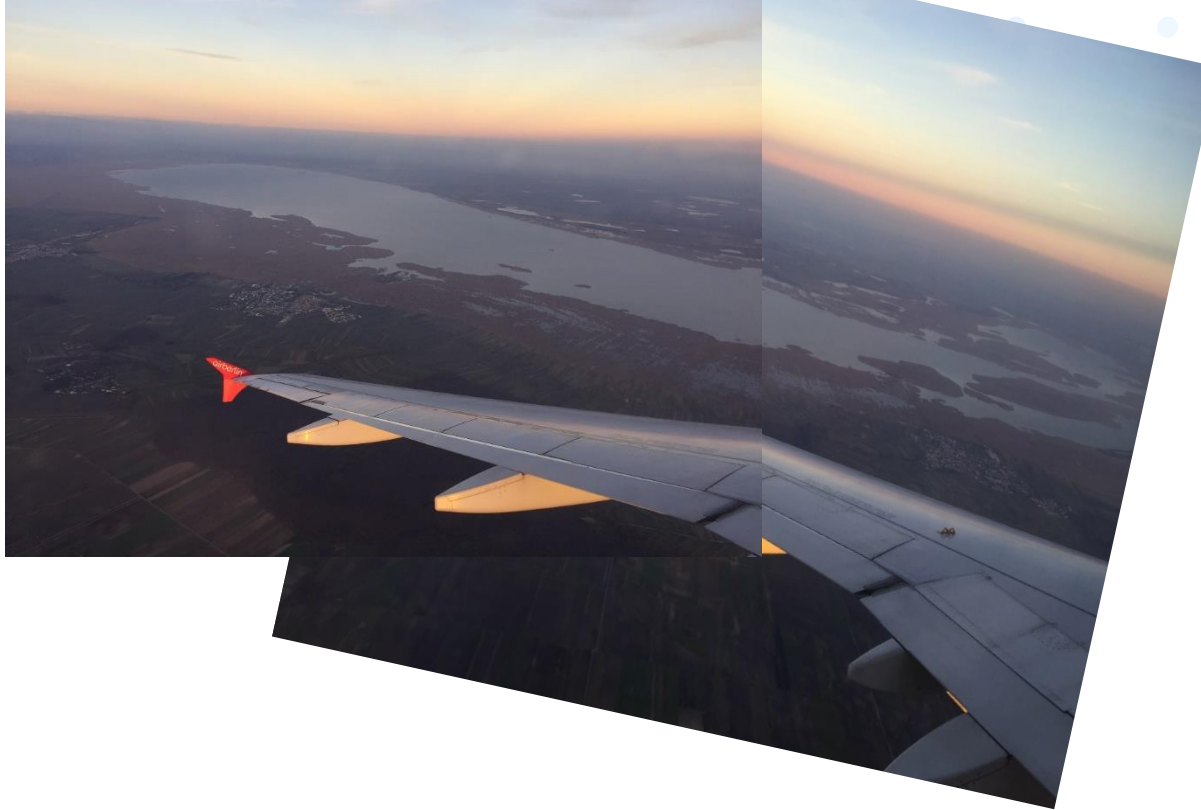
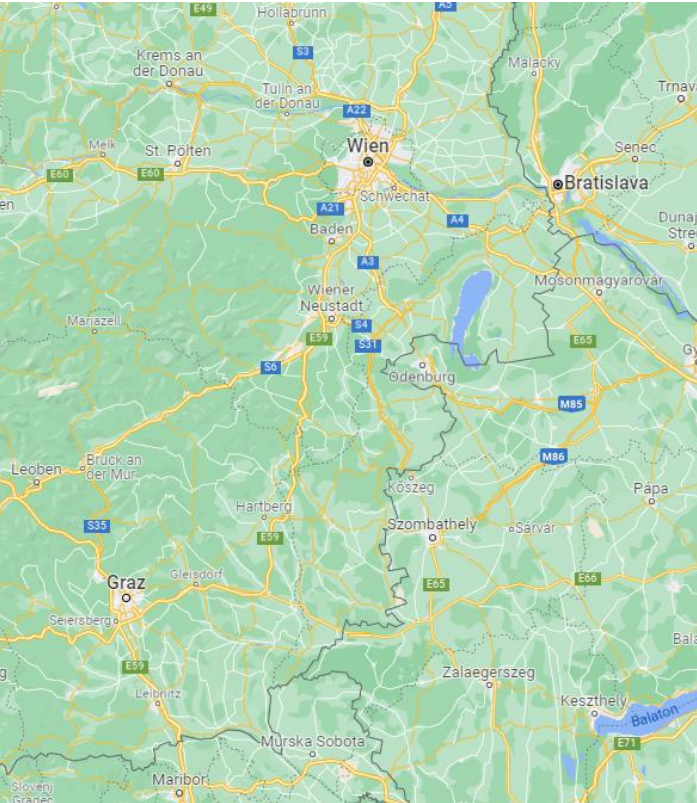
SAS Software D-A-CH

[Link](#)

SAS Communities Library
We're smarter together. Learn from this collection of

[Artikelsammlung
auf SAS Communities](#)

Lake Neusiedl Burgenland, Austria 47°50'N 16°45'E



Historischer Tiefststand am Neusiedler See

burgenland ORF.at

Burgenland-News Magazin Landesstudio Volksgruppen Ganz Österreich



POLITIK

Land plant Wasserzufuhr für Neusiedler See

Der Wasserstand des Neusiedler Sees ist derzeit bedenklich niedrig. Laut Experten des Landes für Wasserwirtschaft könnte der See in absehbarer Zeit sogar komplett austrocknen. Für den zuständigen SPÖ-Landesrat Heinrich Dörner führt an einer technischen Wasserzufuhr kein Weg vorbei.

29. Mai 2020, 13:30 Uhr

[f](#) [t](#) [v](#)

burgenland ORF.at

Burgenland-News Magazin Landesstudio Volksgruppen Ganz Österreich



CHRONIK

Neusiedler See: Historischer Tiefststand für Mai

Im Mai ist der Pegelstand des Neusiedler Sees seit Beginn der Aufzeichnungen im Jahr 1965 noch nie so niedrig wie heuer gewesen. Normalerweise ist der Wasserstand im Frühling noch höher und sinkt über den Sommer. Die Segelboote haben heuer aber schon jetzt Probleme.

21. Mai 2020 / BURGENLAND 619

Neusiedler See trocknet zunehmend aus

Der mittlere Wasserstand liegt unter dem langjährigen Tiefstwert. Nun wird wieder über eine Wasserzufuhr diskutiert



29. Mai 2020 / NEUSIEDLER SEE 333

Grüne und WWF gegen Wasserzufuhr in den Neusiedler See

Landtagsabgeordneter Spitzmüller spricht von einer Gefahr für ein sensibles Natursystem, der WWF gar eine ökologische Katastrophe



26. Mai 2020 / KLIMASTATUSBERICHT 823

Wie die Klimakrise Österreich erfasst: Dürre, Hitze und Starkregen häufen sich

Hitzesommer und Rekordniederschläge sind längst keine Ausnahme mehr. Künftig werden lange und extreme Wetterperioden laut Klimaforschern zunehmen



K Wien

Seite 18 | Sonntag, 17. Mai 2020 | KLIMASTATUSBERICHT

„Verdammt wenig Wasser“

Zu wenig Regen. Ist das Wasser weg, ist der Tourismus weg. Der Neusiedler See leidet unter der Trockenheit. Nicht nur er

Palast 90
Rampel
Der Neusiedler See
gibt sich nicht so
aufgebläht wie im
Sommer. Die
Wasserwirtschaft
denk an einen
Eimer mit dem See

1865
Im Jahr
1865 gab es
den ersten
Wasserstand
im Neusiedler
See. Er war
höher als
heute

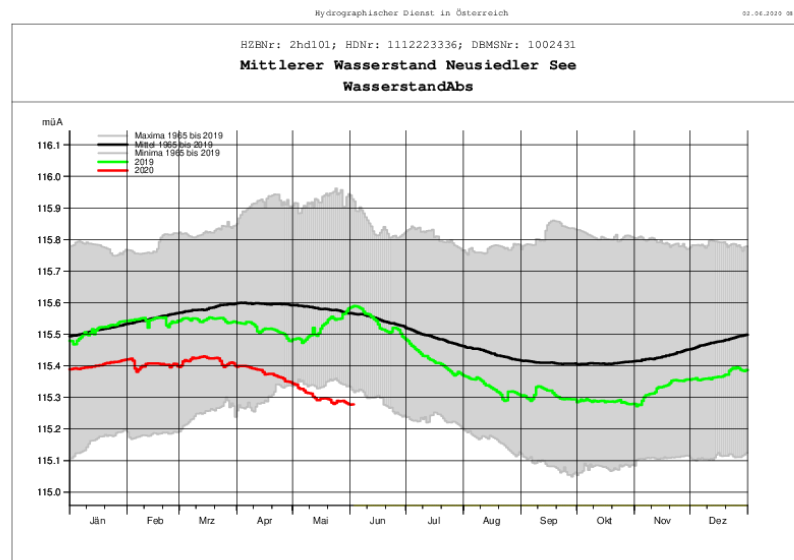
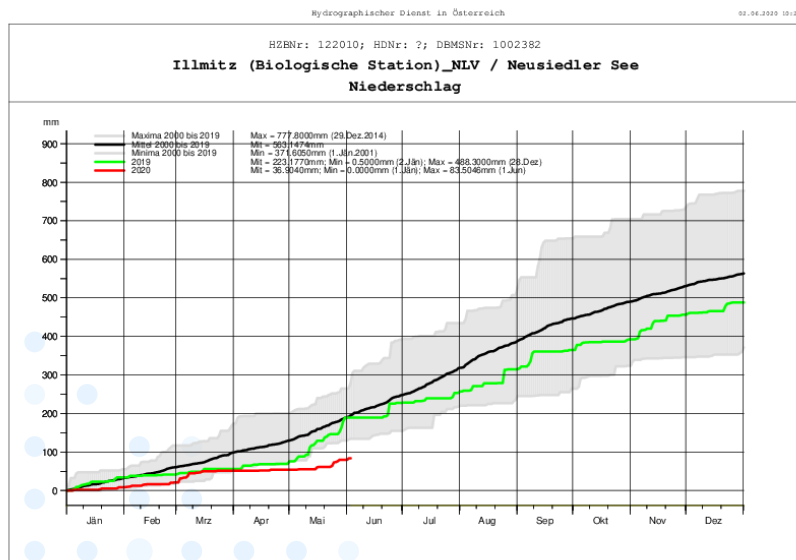
1-2
Klimaforscher
sagen: Die
Wetterlagen
werden
noch extremer
werden. Das
ist ein
Warnzeichen

Wie read out der gegen Neusiedler See, auch die der starkige Wasserstand auch in Brandenfeld befindet. Dürre, Hitze, Wind, weniger - in der Nacht



Wie kommt das?

- 80% -90% des Wasserhaushalts wird durch Regenwasser gespeist
- Jänner bis Mai 2020 waren extrem trocken
- Karl Maracek (Hydrologie Burgenland): Niederschlag im Winter fördert den Aufbau von Wasserreserven (nur geringe Verdunstung)



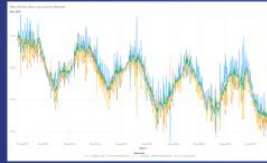
Wie ist der Autor involviert?



Analyse der hydrologischen Daten zum Neusiedler See im Rahmen meiner Lehrveranstaltung an der FH-Burgenland

Task 1.1

Erstellen eines Liniendiagramms pro Station



Rollenzuordnungen
Filtern auf Wasserstand (und Messstellen)
Zuweisung von Farben zu Messstellen
Override System-Limit for #DataPoints



Task 3.4

Create a cumulative chart per year for the number of hours > 30°C



Similar to task 3.1
Be careful with the definition
GREATER THAN or GREATER EQUAL

SAS

Task 4.3

Detail analysis of the scatter plot



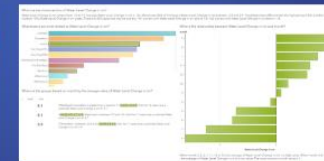
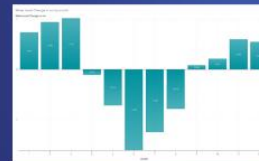
Subgroup the dots by general (average) water level

SAS

Task 5.2

Study Individual Variables and their Influence

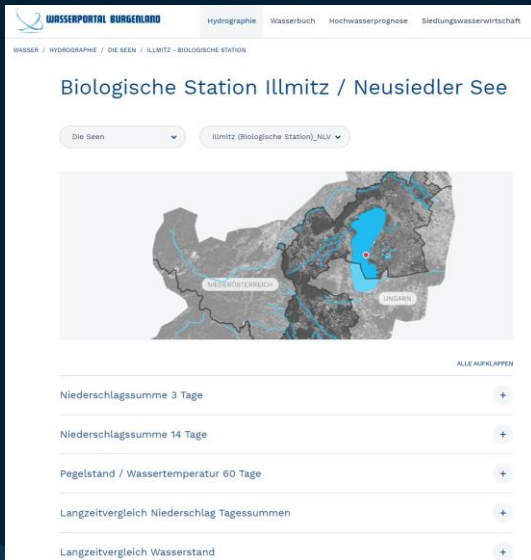
Use Task „Automated Analysis“



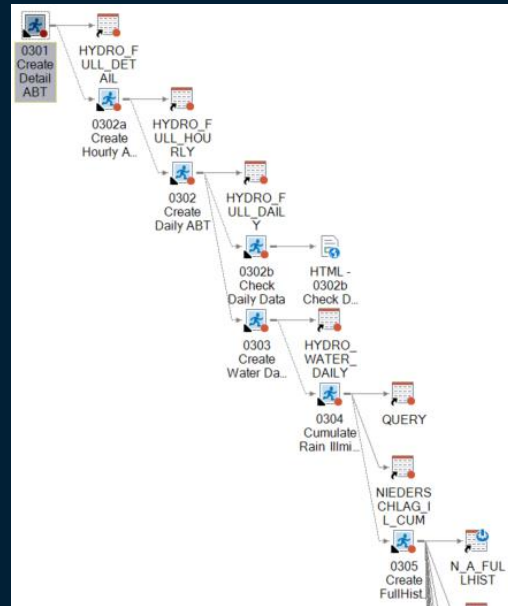
SAS

Überblick über die Vorgangsweise

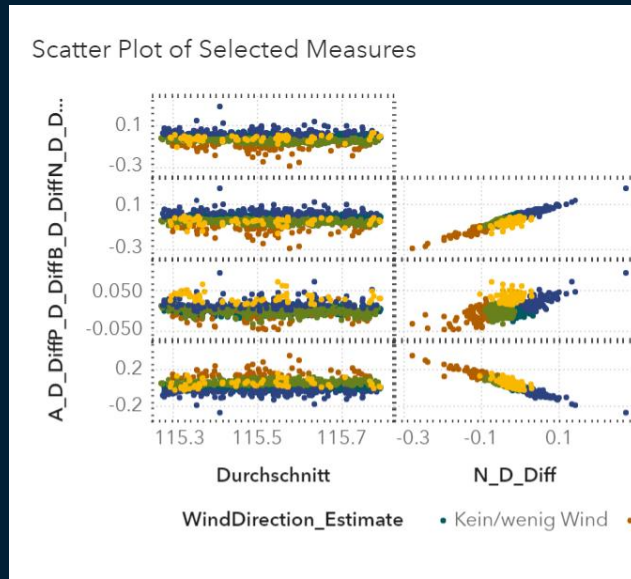
Laden der Daten von der Hydrologie Burgenland



Verknüpfen, Aggregieren
Plausibilitätsprüfung, Aufbereiten



Deskriptive Statistiken, Kausalitäts-
analysen, Simulationen





Langzeitvergleich Niederschlag Monatssummen

Standort

Bilder

```

31.08.2021 21:00:00 115,25
31.08.2021 21:15:00 115,25
31.08.2021 21:30:00 115,24
31.08.2021 21:45:00 115,23
31.08.2021 22:00:00 115,22
31.08.2021 22:15:00 115,24
31.08.2021 22:30:00 115,23
31.08.2021 22:45:00 115,24
31.08.2021 23:00:00 115,23
31.08.2021 23:15:00 115,24
31.08.2021 23:30:00 115,22
31.08.2021 23:45:00 115,23
01.09.2021 00:00:00 115,24

```

```

filename w_n url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_neusiedl_6.txt" ; *** Neusiedl;
filename w_b url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_breitenbrunn_16.txt" ; *** Breitenbrunn;
filename w_p url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_podersdorf_5.txt" ; *** Podersdorf;
filename w_r url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_rust_7.txt" ; *** Rust;
filename w_m url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_moerbisch_8.txt" ; *** Mörbisch;
filename w_i url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_illmitz_4.txt" ; *** Illmitz;
filename w_a url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_apetlona79_3.txt" ; *** Apetlon "Nuller";
filename w_d url "http://wasser.bgld.gv.at/fileadmin/hydromap/pegelstationen/datenbereitstellung/W_s_mwstneusee_15.txt" ; *** Durchschnitt;

data hydro1.w_n_20210930;
filename n_i url "http format OrtID $1. MessID $1. DatumZeit datetime18. Wert 8.2; xt" ; *** Illmitz - Nieders
filename n_a url "http *** Referenz zum Filename Link auf die Homepage; 3.txt" ; *** Apetlon - Nieders
filename t_i url "http infile w_n; xt" ; *** Illmitz - Tempera
*** Lese die enthaltenen Variablen;
input @1 datum ddmmyy10. @12 Zeit hhmmss. Wert2 $ ;
*** Extrahiere Datum, Zeit, Ort, Mess-ID, Wert in eine Variable;
DatumZeit = dhms(datum, hour(zeit), minute(zeit), second(zeit));
OrtID = upcase(substr("w_n", 3, 1));
MessID = upcase(substr("w_n", 1, 1));
Wert = input(tranwrđ(Wert2, ',', '.'), 8.2);
drop Wert2 datum zeit;
run;

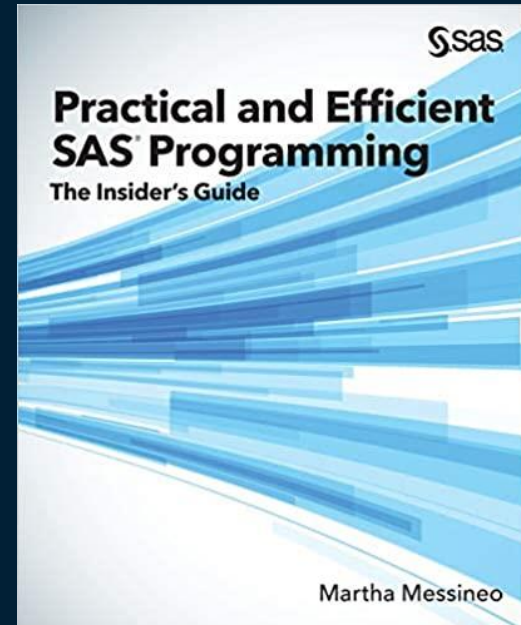
```

Programmier-Tipp für die DHMS Funktion

```
DatumZeit = dhms (datum, hour (zeit) , minute (zeit) , second (zeit) ) ;
```

Oder besser

```
DatumZeit = dhms (datum, 0, 0, zeit) ;
```



Wie wollen wir die Daten anordnen?

Wasser - Apetlon

Wasser - Illmitz

Niederschlag - Illmitz

Wasser - Neusiedl

Temperatur - Illmitz

(LONG) Longitudinal, Transactional

Wasser - Apetlon

Wasser - Illmitz

Wasser - Neusiedl

Temperatur - Illmitz

Niederschlag - Illmitz

Wie sieht unsere Datenstruktur (z.B für Korrelationsanalysen) aus? – WIDE Format

Wasser - Neusiedl

Wasser - Illmitz

Wasser - Apetlon

Temperatur - Illmitz

Niederschlag - Illmitz

Wie sieht unsere Datenstruktur aus?

Mess-Stations Daten	Mess-Größen Daten	Wasser - Neusiedl
Mess-Stations Daten	Mess-Größen Daten	Wasser - Illmitz
Mess-Stations Daten	Mess-Größen Daten	Wasser - Rust
Mess-Stations Daten	Mess-Größen Daten	Temperatur - Illmitz
Mess-Stations Daten	Mess-Größen Daten	Niederschlag - Illmitz

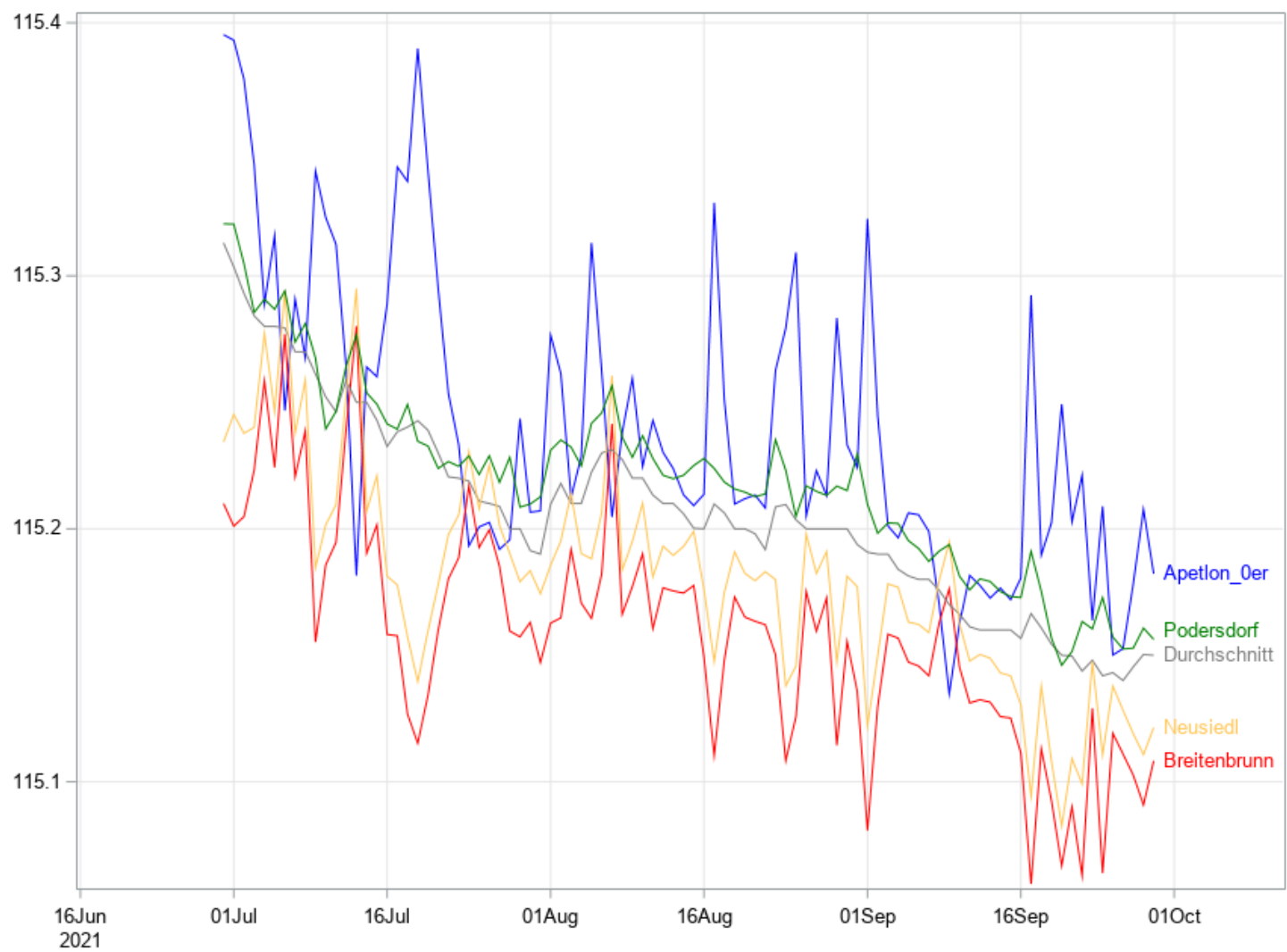
Aggregations-Level ?

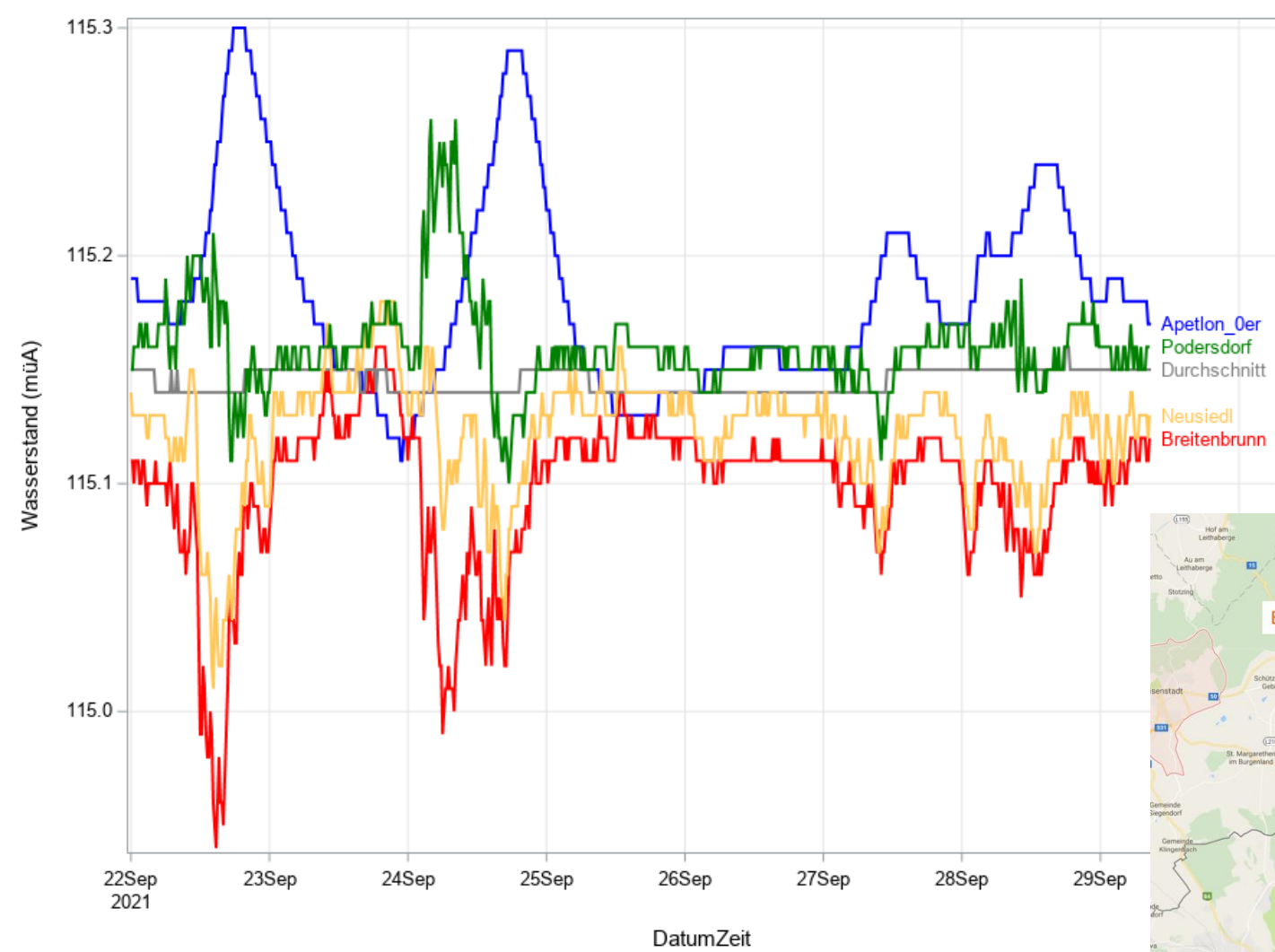
15 min

Stündlich

Täglich

Monat





Make sure, that the same group always has the same color.

```
ods graphics / reset width=24cm height=18cm imagemap;
```

```
%let linethickness = 2;
```

```
data Attr_Messstelle_Color;
```

```
format id $10. value $13. LINECOLOR $12. thickness $2.;
```

```
id="Messstelle"; value="Apetlon_0er"; LINECOLOR="blue";
```

```
id="Messstelle"; value="Neusiedl"; LINECOLOR="lightorange";
```

```
id="Messstelle"; value="Breitenbrunn"; LINECOLOR="red";
```

```
id="Messstelle"; value="Podersdorf"; LINECOLOR="green";
```

```
id="Messstelle"; value="Durchschnitt"; LINECOLOR="grey";
```

```
;
```

```
run;
```

**Specify the LINECOLOR
(and other options)
in an ATTRIBUTE dataset.**

```
linethickness=&linethickness.; output;
```

```
linethickness=&linethickness.; output;
```

```
linethickness=&linethickness.; output;
```

```
linethickness=&linethickness.; output;
```

```
linethickness=&linethickness.; output;
```

```
proc sgplot data=HYDRO_FULL_DETAIL (where=(datepart(datumzeit) >= today()-7
```

```
and substr(messstelle, 1, 1) not in ("M", "R", "I")
```

```
and substr(messgroesse, 1, 1) in ("W"))) )
```

```
dat|trmap=Attr_Messstelle_Color;
```

```
series x=DatumZeit y=Wert / group=Messstelle curvelabel curvelabelpos=max attrid=Messstelle ;
```

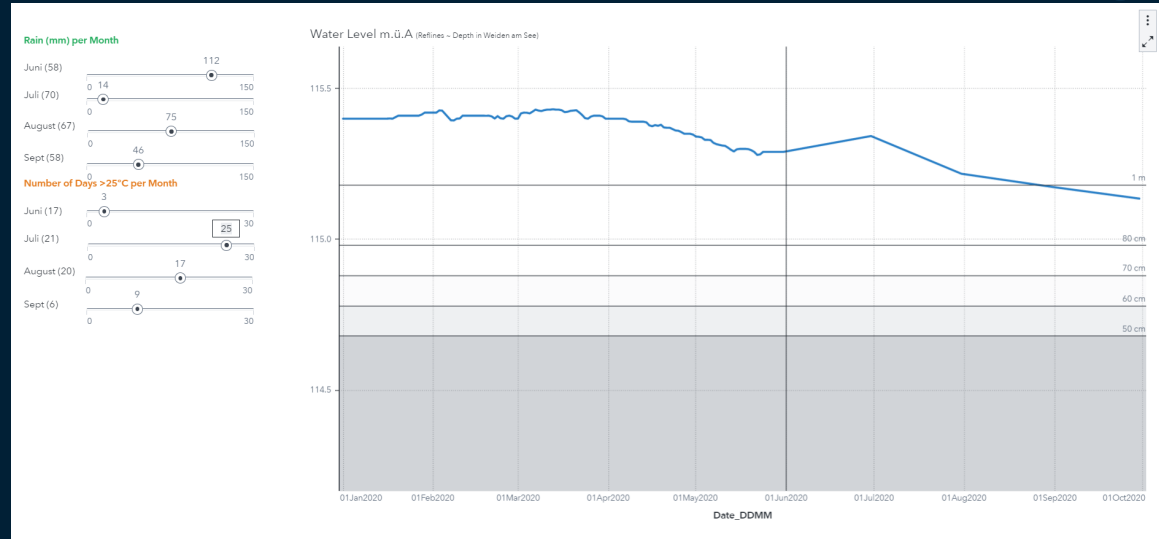
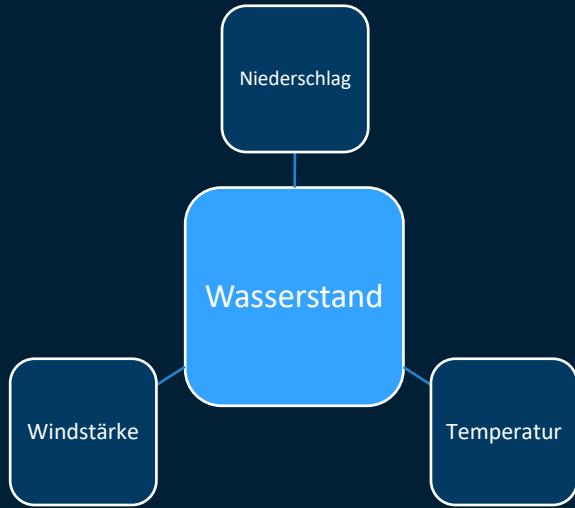
```
xaxis grid;
```

```
yaxis label= "Wasserstand (müA)" grid;
```

```
run;
```

```
ods graphics / reset;
```

Use a regression model do explain and quantify the relationship between different factors



- Data Preparation + Feature Engineering
- Modeling with the GLMSELECT procedure
- Scenario Analysis in SAS Visual Analytics

Daily data in a WIDE format

Wasser - Neusiedl

Wasser - Illmitz

Wasser - Apetlon

Temperatur - Illmitz

Niederschlag - Illmitz

Feature Engineering

```
*** Temperature;
Temp_GT30 = (TempMax > 30);
Temp_GT25 = (TempMax > 25);

*** Days Since Last Rain;
retain DaysSinceLastRain 0;
if RainSum > 0 then DaysSinceLastRain = 0;
else DaysSinceLastRain + 1;

*** Water Shift between North/South, East/West;
WShift_N_A = Water_N - Water_A;
WShift_B_P = Water_B - Water_P;
WShift_R_P = Water_R - Water_P;
WShift_B_I = Water_B - Water_I;
```

Aggregate from daily to a monthly level

```
select intnx('month',date,0,'Begin') as YMMM format = yymm7.,  
       month as month,  
       mean(WaterLevel) as WaterLevelMean format=8.2,  
       sum(WaterLevel*(day(date)=1)) as WaterLevel_lofMonth format =8.2,  
       sum(RainSum) as RainSum format = 8.,  
       sum(RainSum > 0) as Cnt_RainDays,  
       max(DaysSinceLastRain) as MaxDaysSinceLastRain,  
       mean(TempMean) as TempMean format = 8.1,  
       sum(Temp_GT30) as Cnt_TmpGT30 ,  
       sum(Temp_GT25) as Cnt_TmpGT25,  
       max(WShift_N_A) as WShift_N_A format=8.2,  
       max(WShift_B_P) as WShift_B_P format=8.2,  
       max(WShift_R_P) as WShift_R_P format=8.2,  
       max(WShift_B_I) as WShift_B_I format=8.2
```

Observe which
aggregation
statistic you should
use for each
feature!

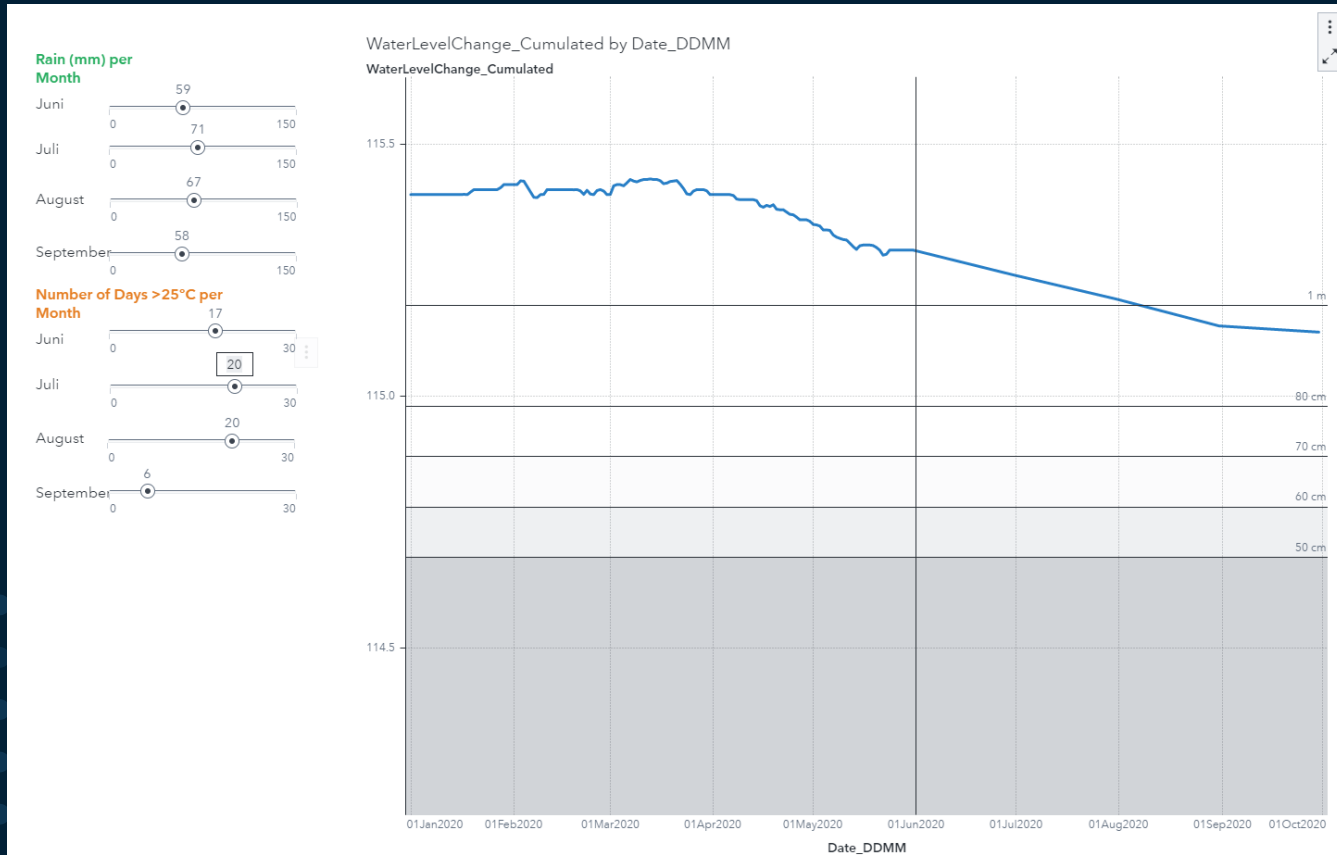
Use SAS/STAT (or SAS Visual Statistics) to train the model

```
proc glmselect data=monthly_abt_month_sort;  
  model WaterLevelChange =  
    WaterLevelMean RainSum Cnt_Raindays  
    MaxDaysSinceLastRain  
    TempMean Cnt_TmpGT30 Cnt_TmpGT25  
    WShift_N_A WShift_B_P WShift_R_P Shift_B_I  
  /selection=backward;  
  where month in (6,7,8,9) and year LE 2019;  
  code file="&path.\Hydro_WaterChange_Mod1.0.sas";  
run;
```

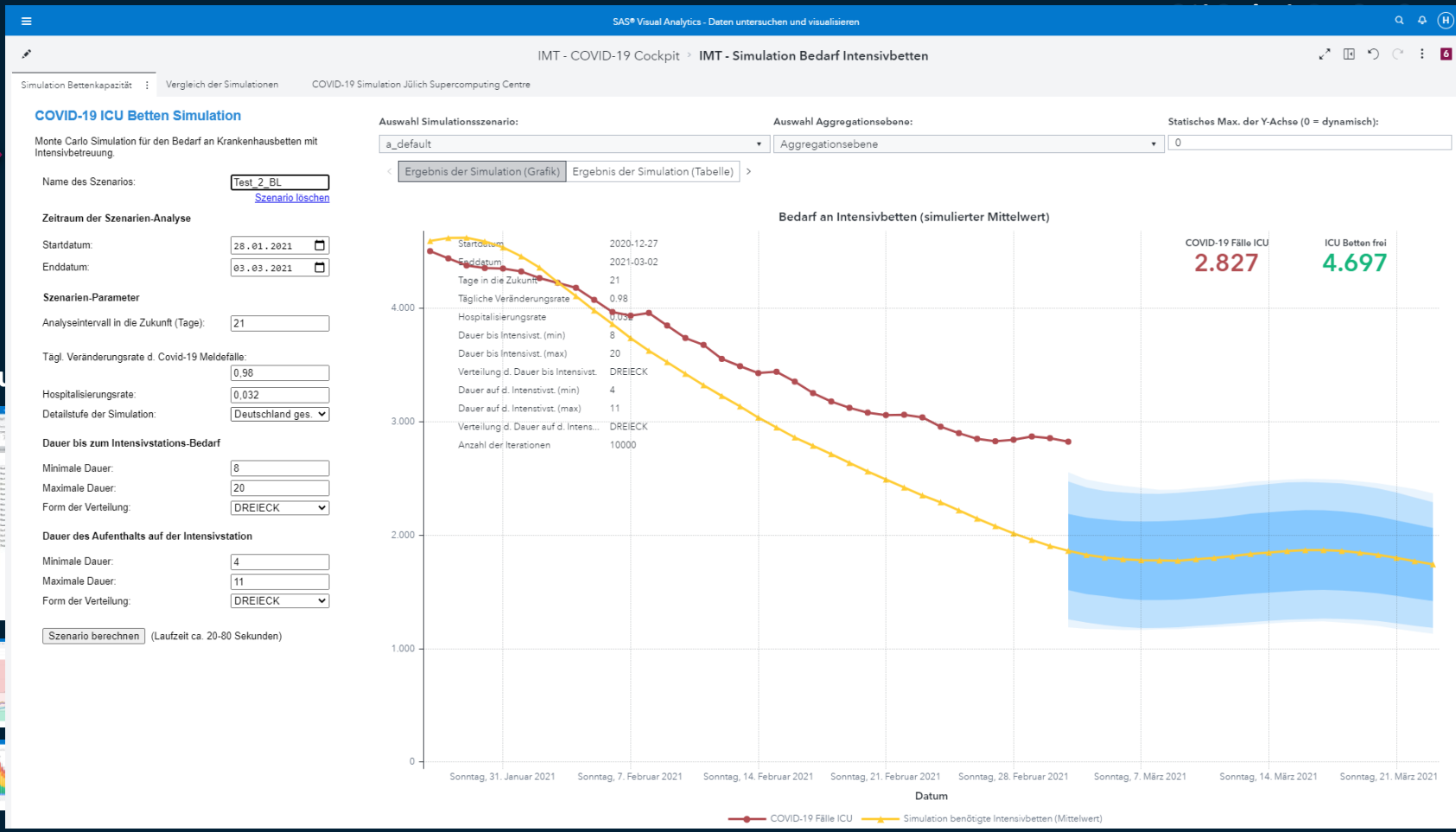
Parameter Estimates					
Parameter	DF	Estimate	Standard Error	t Value	Pr > t
Intercept	1	-58.153454	7.587593	-7.66	<.0001
RainSum	1	1.096672	0.062758	17.47	<.0001
Cnt_TmpGT25	1	-3.330063	0.338838	-9.83	<.0001

- ← Average monthly water loss per summer month
- ← Rain adds to the water level with a factor of ~ 1
- ← Day >25 °C „costs“ 3.3 mm of water level

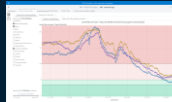
Using SAS Visual Analytics to interactively calculate prediction for different scenarios



Information Management Tool des BMG



Aktu

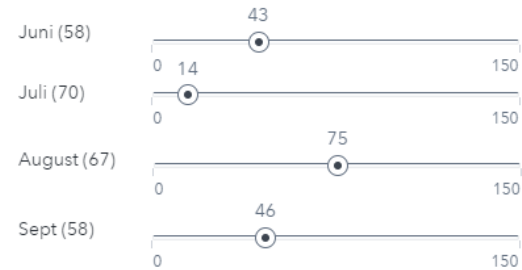


Pre-Calculate outcomes for all possible scenarios and filter according to the selected values

```
Data hydro3.PredWaterLevelChange;
format month 8. P_WaterLevelChange 8.2;
do month = 5 to 10;
  do Cnt_TmpGT25 = 0 to 30;
    do RainSum = 0 to 200 by 5;
      %include "\Hydro_WaterChange_Mod1.0.sas";
      output;
    end;
  end;
end;
run;
```

month	RainSum	Cnt_TmpGT25	P_WaterLevelChange
6	60	23	-68.94
6	155	14	65.21
6	160	23	40.72
6	125	5	62.28
6	50	9	-33.29

Rain (mm) per Month



Number of Days >25°C per Month



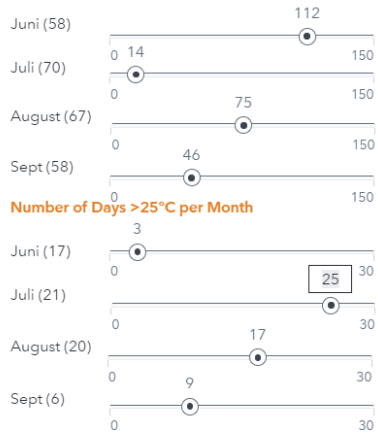
Usage function “Cumulative Period” in SAS Visual Analytics to plot the course over 4 months

The screenshot displays the SAS Visual Analytics configuration interface for a function named "CumulativePeriod".

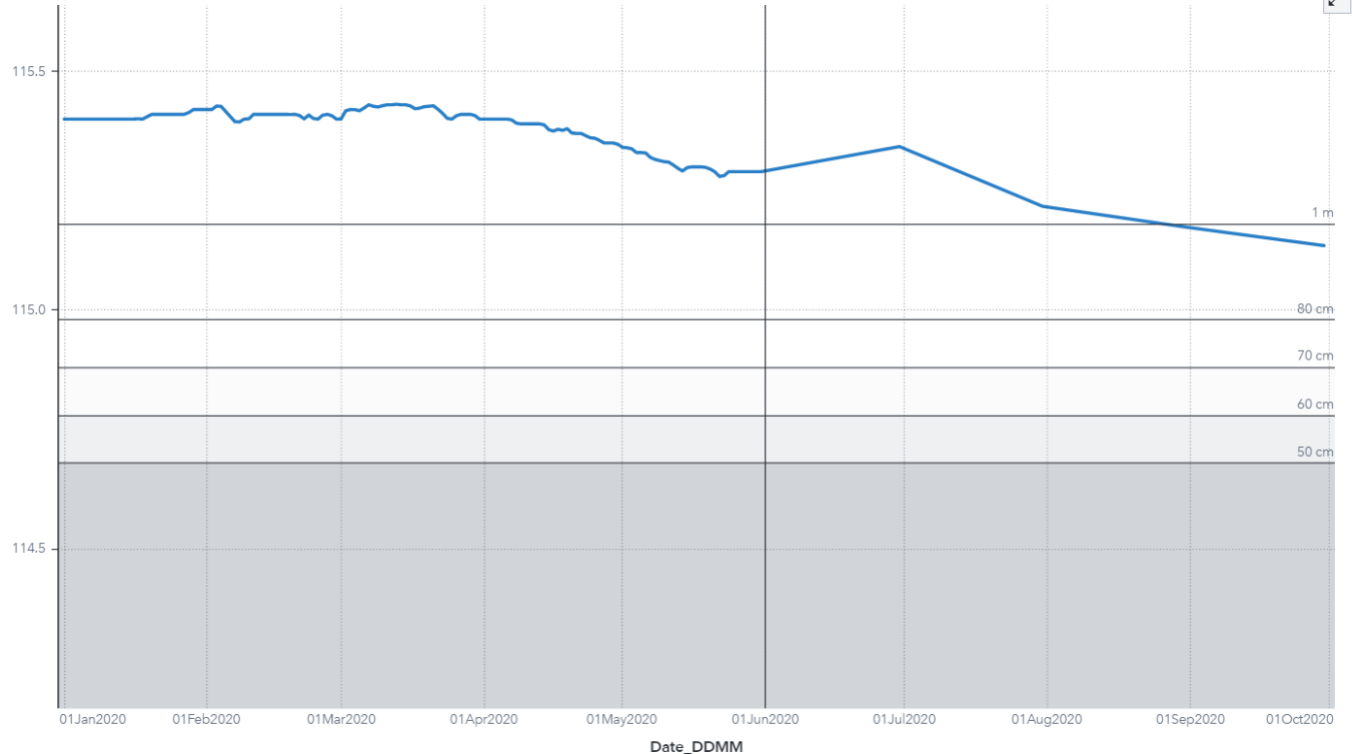
- Name:** WaterLevelChange_Cumulated
- Result Type:** Aggregated Measure
- Base expression:** P_WaterLevelChange_Daily
- Visual/Text:** Visual (selected)
- Configuration:**
 - Function: CumulativePeriod
 - IgnoreAllTimeFra... (dropdown)
 - Date_DDMM (dropdown)
 - Inferred_ (dropdown)
 - Inferred_ (dropdown)
 - Start: 0 (input field)
 - End: 1 (input field)
 - Full_ (dropdown)
 - No selection (input field)
- Messages:** (0)

Using SAS Visual Analytics to interactively calculate prediction for different scenarios

Rain (mm) per Month



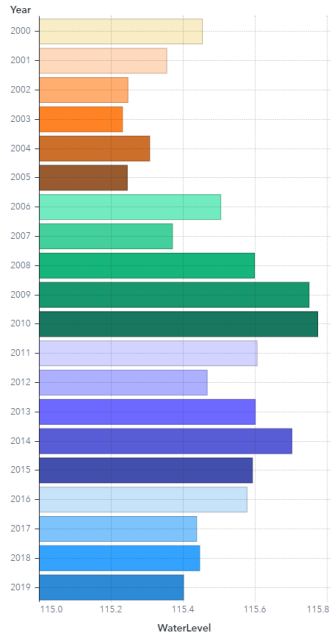
Water Level m.ü.A (Refines - Depth in Weiden am See)



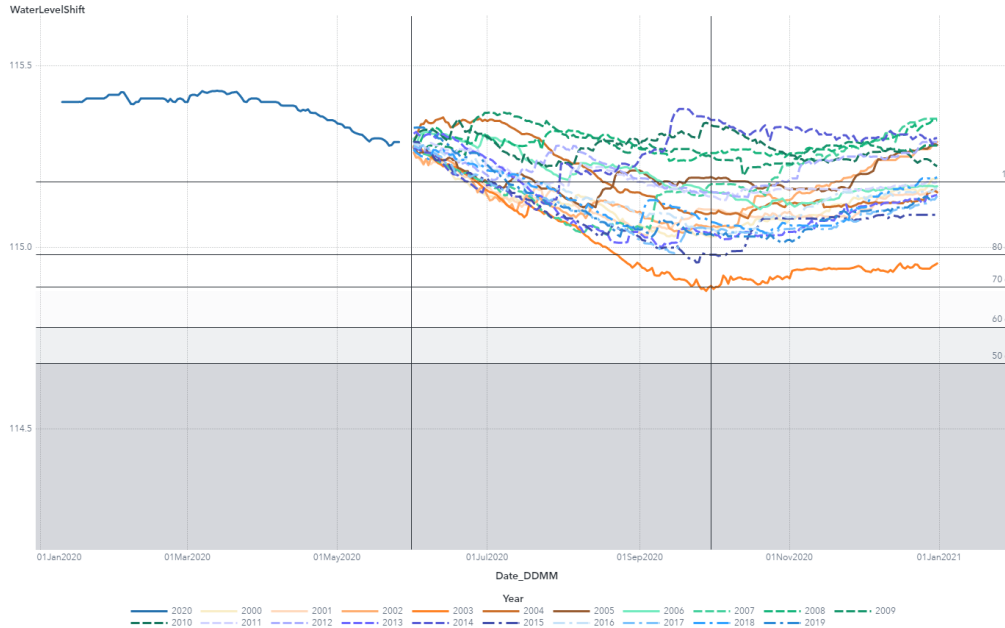
Was wäre wenn, das Wetter im Sommerhalbjahr so ist, wie ...

- im extrem trockenen Jahr 2003,
- im niederschlagsreichen Jahr 2014,
- ...

Average Water Level June-September



Water Level müA (Refines = Depth at Weiden am See)



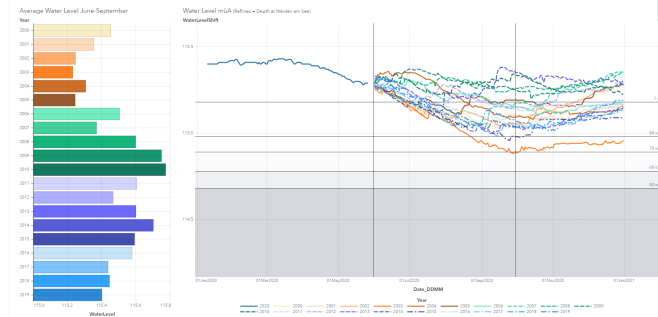
How to create such a chart:

1. Create a date axis only with DD.MM (ignore YY)

```
format DayMonth2018 date9.;
```

```
DayMonth2018 = mdy(month(datum), day(datum), 2018);
```

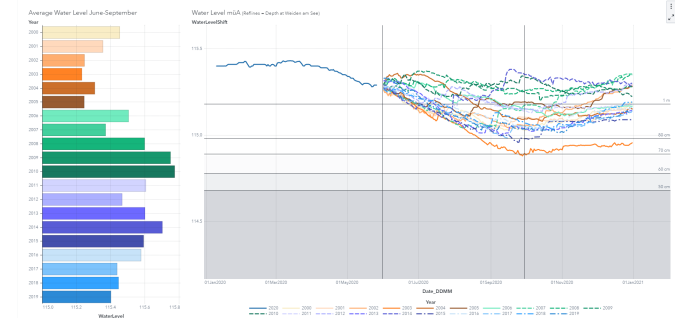
2. Plot the actual year until the last available day
3. For the next day: Shift the values from the historic years to the value of the last available day (*“as if the year would continue from today as a certain historic year”*)
4. Plot this in a line chart with YEAR as BY-GROUP



```

proc sql noprint;
*** Level at CutDate;
select WaterLevel
into :WaterLevel_AtCutDate
from hydro3.Daily_ABT_2020
where date=&CutDate.;
*** Level at CutDate DDMM per year;
create table year_diff_at_cutdate
as
select year,
       WaterLevel - &WaterLevel_AtCutDate. as Yearly_Diff_CutDate format = 8.3
from hydro3.Daily_ABT_2020
where date_ddmm = &CutDate.+1;
quit;

```



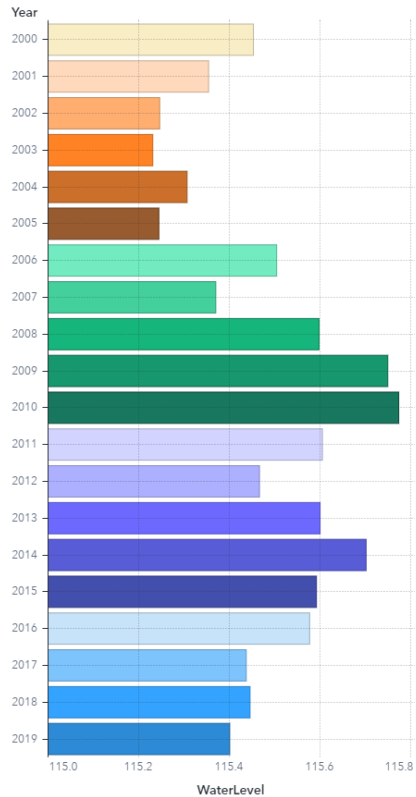
```

proc sql;
create table hydro3.Water_Level_Concat
as select
  a.year
,a.date
,a.date_ddmm
,a.WaterLevel
,case when (a.date_ddmm > &CutDate.) then a.WaterLevel - b.Yearly_Diff_CutDate
      else a.WaterLevel end
      as WaterLevelShift format = 8.2
from hydro3.Daily_ABT_2020 as a
left join work.year_diff_at_cutdate as b
on a.year = b.year
order by date
;
quit;

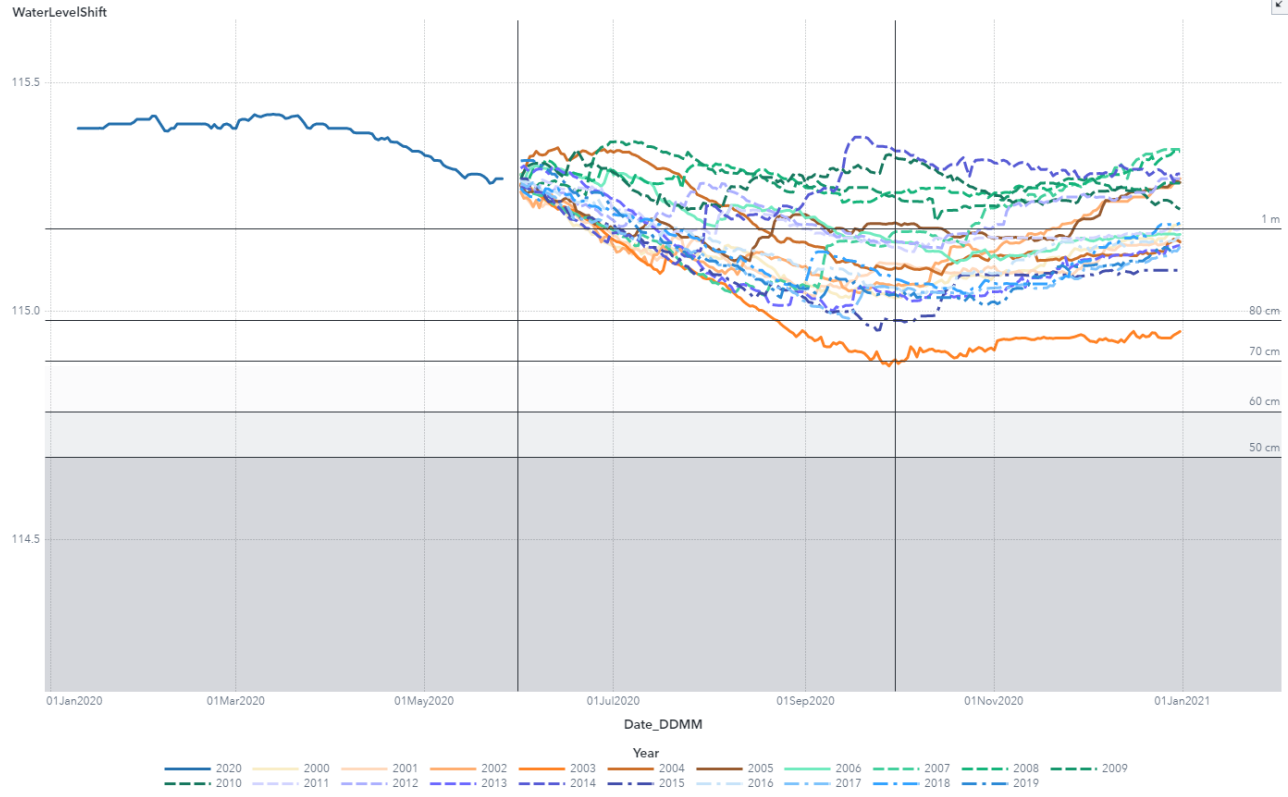
```

Interactive Display in SAS Visual Analytics: Selecting a bar on the left filters line chart on the right

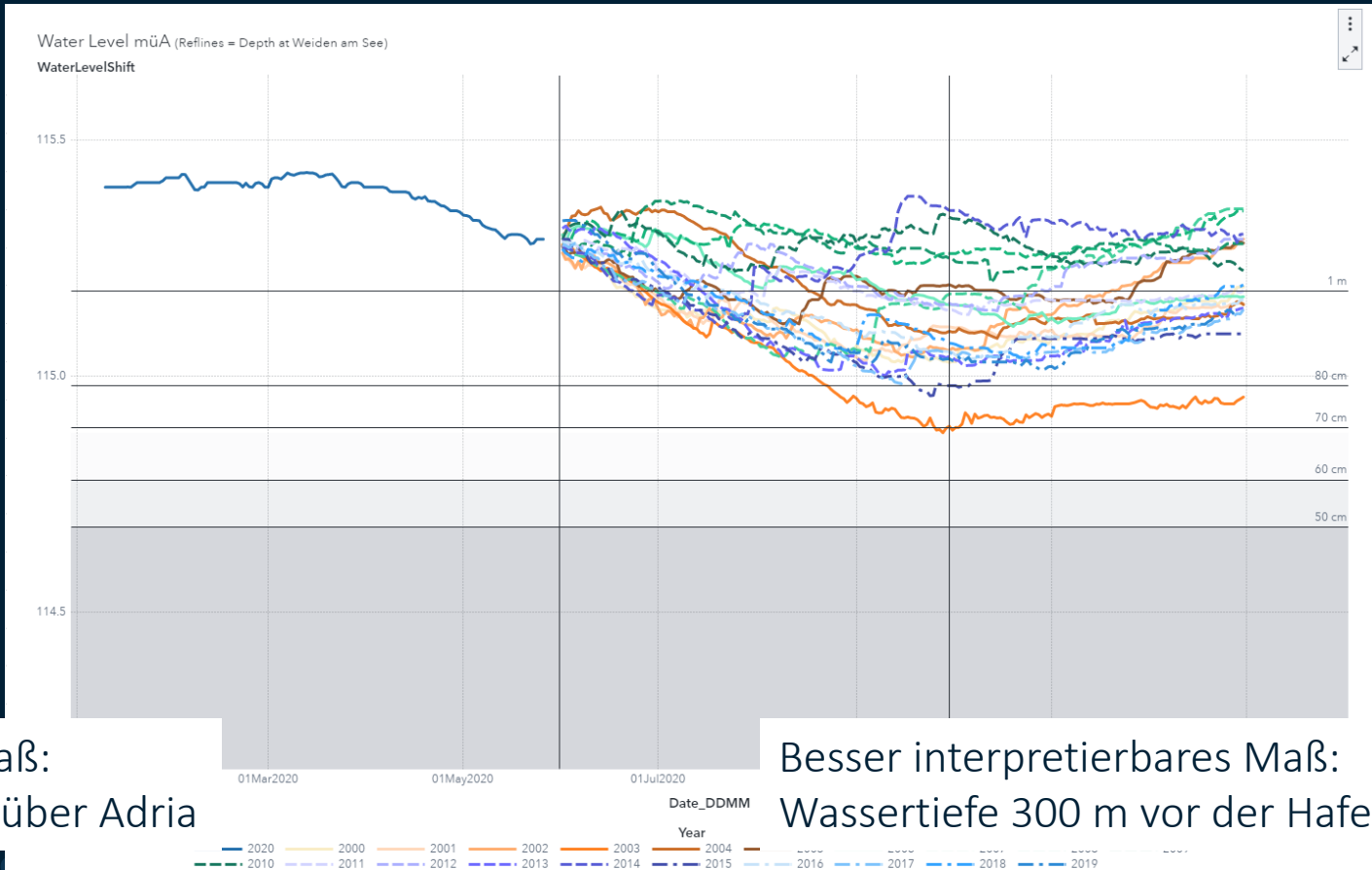
Average Water Level June-September



Water Level müA (Refines = Depth at Weiden am See)



Display Tipp: Use interpretable scales at your graph axes



Zusammenfassung

- Die Kombination von SAS Datenmanagement und SAS Analytik und SAS Visual Analytics erlaubt es Ihnen, Ihre Analysedaten optisch ansprechend und interaktiv zu präsentieren.
- Die Analyse des Wasserstands ist nur ein Anwendungsbeispiel.
- Diese Methoden sind quer über die Branchen und fachlichen Bereiche im Einsatz.

Data Preparation for Data Science

Data
Assembly

Data Quality
for Analytics

Feature
Generation

Gerhard Svolba,
Data Scientist @SAS
mailto:sastools.by.gerhard@gmx.net

Articles
and Blogs



Webinars



Tipps &
Tricks



Macros &
Downloads

