

# Matematisk modellering

STATISTISK ANALYS

JESPER STOLT

## Innehåll

Förord .....	2
1 - Beskrivning av data .....	3
1.1 – Urval av data .....	3
1.2 – Visualisering av data .....	3
2 – Beskrivande statistik .....	4
2.1 – Tabell över data .....	4
2.2 – Heatmap .....	4
3 – Histogram .....	5
4 – Linjär regression .....	6
4.1 – Visualisering .....	6
5 – Transformerad data .....	7
5.1 – Visualisering .....	7
6 – Residualanalys .....	8
6.1 – Visualisering .....	8
6.2 - Varians .....	8
7 – Sammanfattning .....	9
Referenser .....	10
Appendix 1 – Linjär regression .....	i
Appendix 1.1 — Hallands Väderö A .....	i
Appendix 1.2 – Halmstad flygplats .....	i
Appendix 1.3 – Nidingen A .....	ii
Appendix 2 – Transformerad data .....	iii
2.1 – Hallands Väderö A .....	iii
2.2 – Halmstad flygplats .....	iii
2.3 – Nidingen A .....	iv
Appendix 3 – Residualanalys .....	v
3.1 – Hallands Väderö A .....	v
3.2 – Halmstad flygplats .....	v
3.3 – Nidingen A .....	vi

## Förord

Denna rapport är en del i det avslutande projektet, för kursen Matematisk modellering på Blekinge tekniska högskola. För att skapa grafer och tabeller har JavaScript använts. För att skapa graferna har chartjs-node-canvas använts. För att skapa tabeller har node-html-to-image och för att läsa csv-filerna så har csv-parser använts. För att sedan köra hela analysen har node använts och då har filehandler sparat grafer och tabeller som PNG-filer för att kunna använda i denna rapport.

## 1 - Beskrivning av data

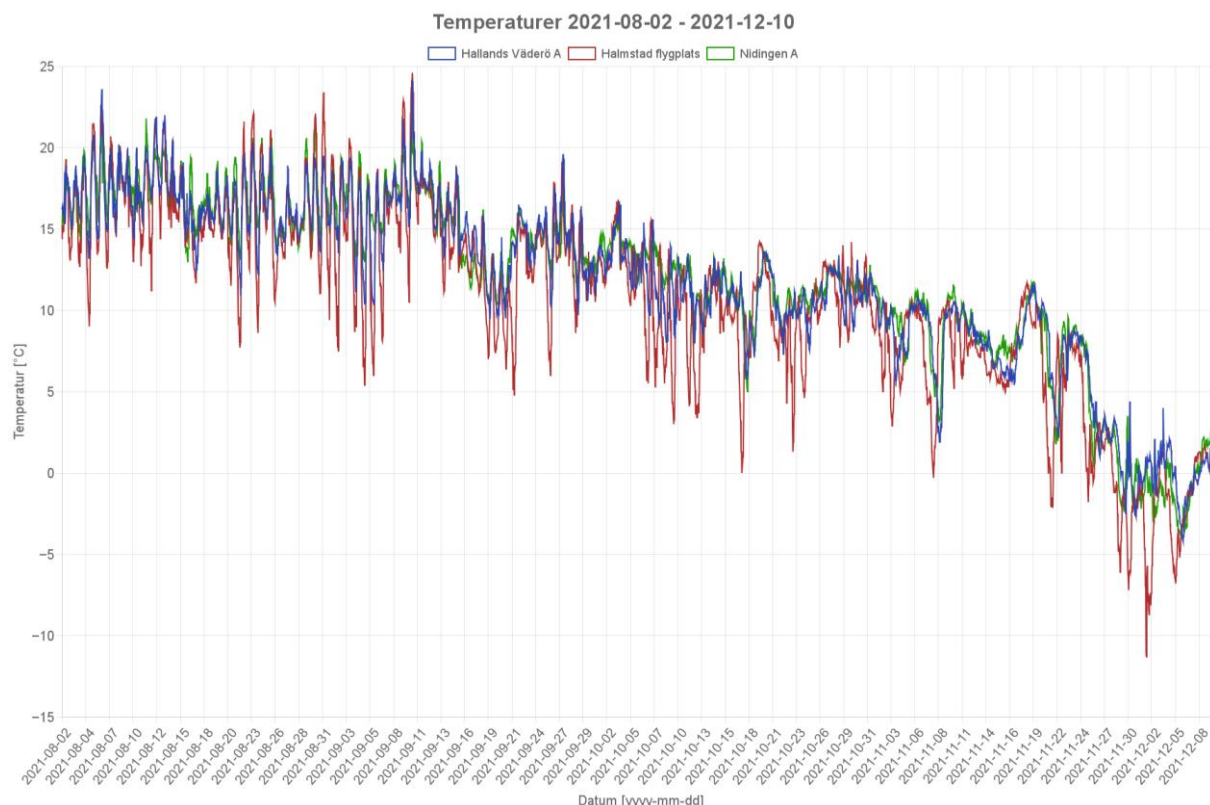
Data för denna rapport är hämtat från SMHI och innefattar lufttemperaturer för tre väderstationer längs västkusten. Temperaturen mäts varje timme och mäts i temperatur Celsius. Data är från fyra månader fram till 2021-12-10. Väderstationerna som använts är Hallands Väderö A [1], som är beläget utanför kusten vid Båstad, vilken mäts 9,172 meter över havet. Halmstad flygplats [2], som är lokaliserat i nordvästra delen av Halmstad, vilket mäts 31 meter över havet. Samt Nidingen A, som finnes utanför kusten mellan Varberg och Kungsbacka, vilket mäts 3,785 meter över havet. Om data för ett mättillfälle saknas är det för att mätare eller station varit ur funktion. Varje mätning har en angivelse för dess kvalitet vilken kan vara antingen G eller Y. Data som har kvalité av typen G är kontrollerad och godkänd, medan data av typen Y Kan vara avvikande eller att den är för ny för att ha kontrollerats.

### 1.1 – Urval av data

Datum	UTC	Lufttemperatur [°C]	Kvalitet
2021-08-02	01:00:00	16.2	G
2021-08-02	02:00:00	16.2	G
2021-08-02	03:00:00	16.4	G
2021-08-02	04:00:00	15.9	G
2021-08-02	05:00:00	16.4	G
2021-08-02	06:00:00	16.4	G

Tabell 1 – Ett urval av data för Hallands Väderö A

### 1.2 – Visualisering av data



Figur 1 – Diagrammet visar en översikt över samtliga mätningar av lufttemperaturen som är gjorda i data. Varje mätstation är visualiserad enligt färgerna ovan. Som rubriken säger sträcker sig mätningarna från 2021-08-02 – 2021-12-10.

## 2 – Beskrivande statistik

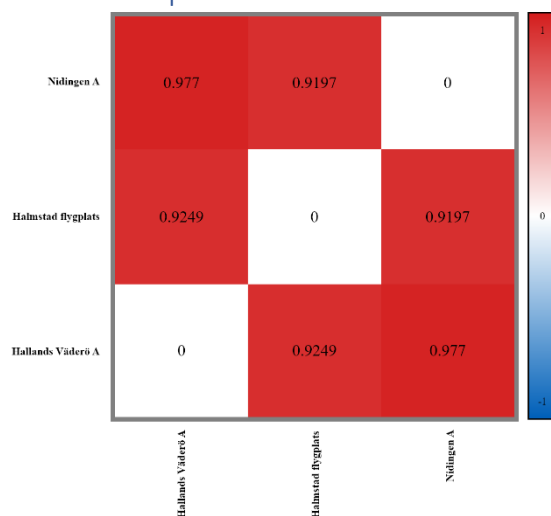
För att beskriva statistiken skapas en tabell med nyckelvärden för datavariablerna. Vi kan se att eftersom samtliga tre väderstationer geografiskt ligger nära varandra, så varierar inte nyckelvärdena nämnvärt mellan stationerna. Vi ser även att det finns starka korrelationer mellan väderstationerna då dess värden ligger nära ett. Vilket även det känns logiskt då vädret mellan olika väderstationer ofta hänger ihop. Ifall det till exempel blir varmare på ett ställe bör det bli varmare på ett anta ställe som är geografiskt nära. Det mest intressanta är att Halmstad flygplats har uppmätt både varmest och kallast väder. Min gissning hade annars varit att Nidingen A hade uppmätt kallast temperatur då den är lokaliserad längst norrut geografiskt. Det kan däremot ha att göra med att Halmstad flygplats mätts på högre höjd än övriga stationer.

### 2.1 – Tabell över data

	Hallands Väderö A	Halmstads flygplats	Nidingen A
Medelvärde [°C]	11.67	10.66	11.89
Standardavvikelse	5.39	6.05	5.5
Minvärde	-4.2	-11.3	-3.8
Maxvärde	24.1	24.6	21.8
Korrelation	HVA-HF: 0.9249	HVA-NA: 0.977	HF-NA: 0.9197

Tabell 2 – En översikt över variablernas nyckeltal.

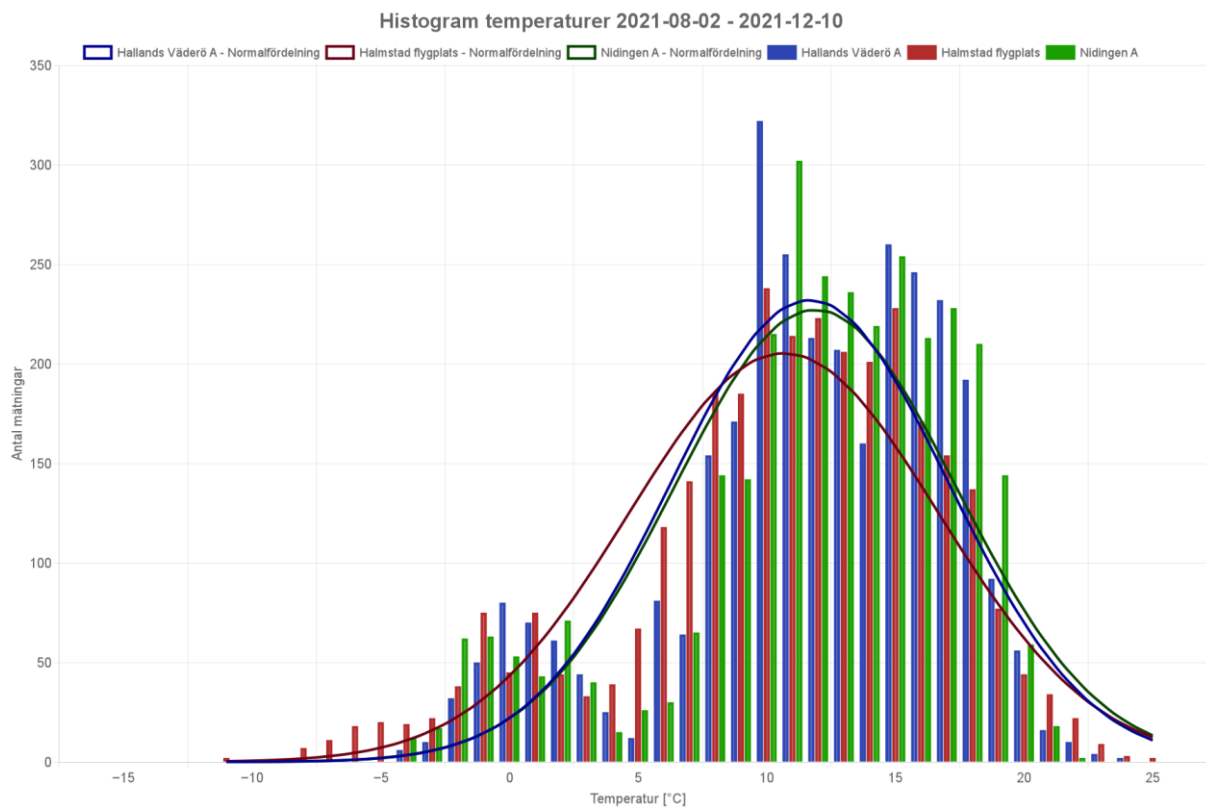
### 2.2 – Heatmap



Figur 2 – korrelationerna mellan respektive variabel.

### 3 – Histogram

För att ytterligare visualisera data inom analys har ett histogram skapats. Histogrammet används för att dela in data i intervall och se hur vanligt förekommande respektive intervall är. För detta histogram har intervall om 1°C använts. Även normalfördelningen för respektive väderstation finns visualiserad i histogrammet.



Figur 3 – Samtliga variablers spridning inklusive deras konfidensintervall.

## 4 – Linjär regression

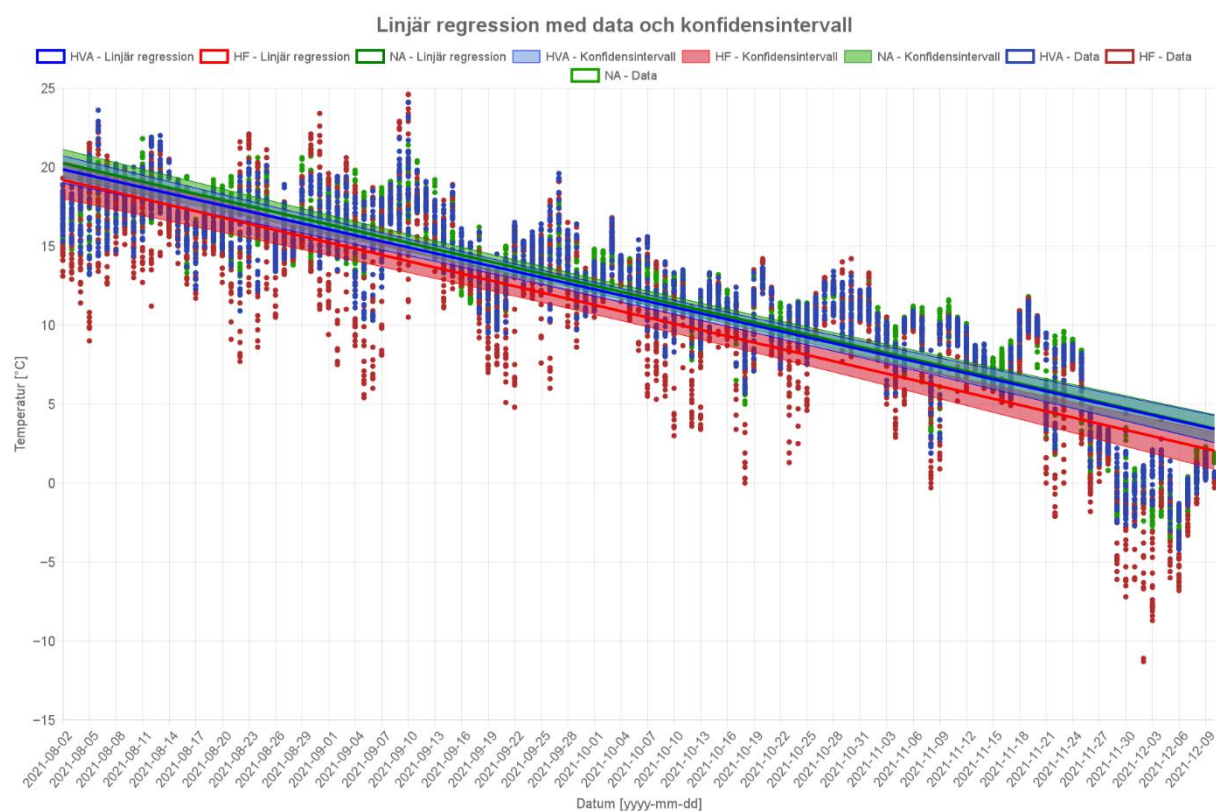
För att visa hur data förändras över tid utförs en linjär analys av data.

	Hallands Värderö A	Halmstads flygplats	Nidingen A
$\alpha$	19.85	19.19	20.25
$\beta$	-0.1264	-0.1319	-0.1291
$y = \alpha + \beta x$	$y = 19.85 - 0.13x$	$y = 19.19 - 0.13x$	$y = 20.25 - 0.13x$
Konfidensintervall $\alpha$	$19.85 \pm 0.18$	$19.19 \pm 0.24$	$20.25 \pm 0.18$
Konfidensintervall $\beta$	$-0.1264 \pm 0.0023$	$-0.1319 \pm 0.0032$	$-0.1291 \pm 0.0024$

Tabell 3 – Visar respektive datavariabels linjära regression.

### 4.1 – Visualisering

Nedan finns en visualisering av alla tre variabler. För att få en bild för respektive variabel enskilt se bilaga 1.



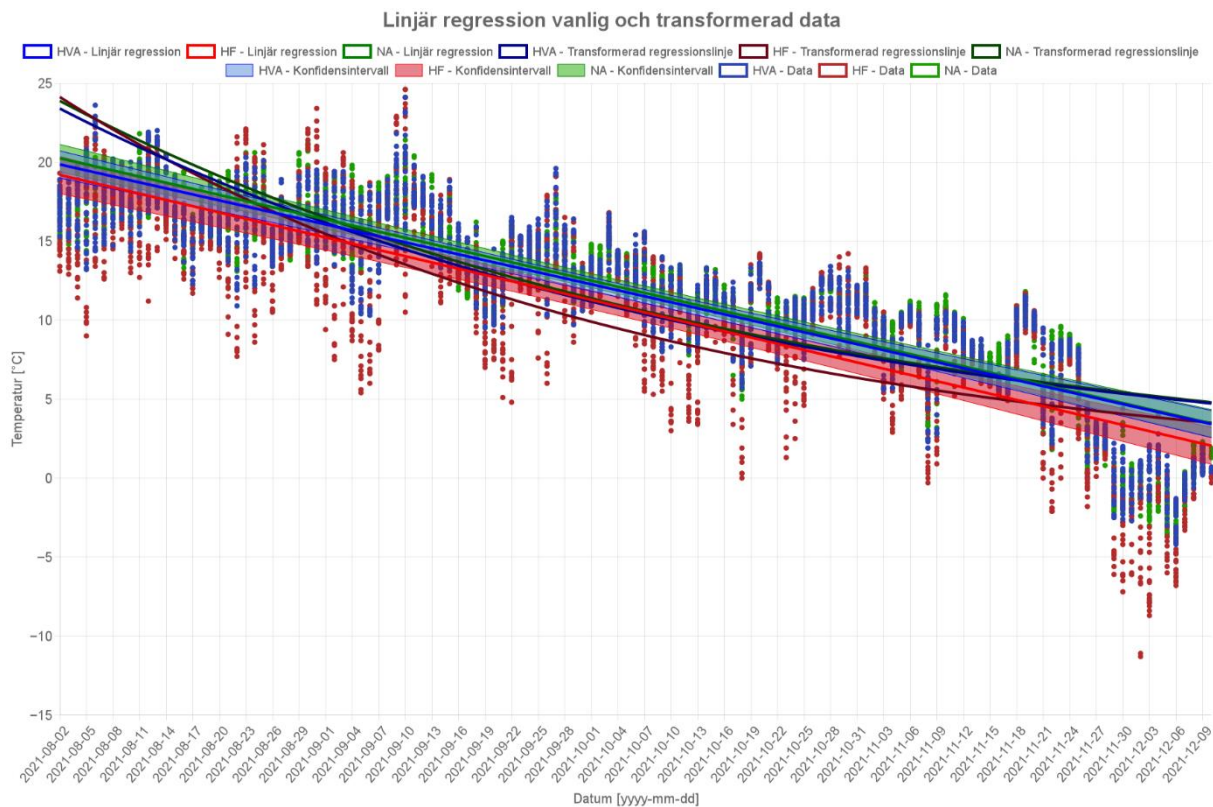
Figur 4 – Vi ser respektive variabels spridning av data samt deras linjära regression med dess konfidensintervall.

## 5 – Transformerad data

Ett ytterligare sätt att utföra en linjär analys på är genom att först transformera data, vilket gör det möjligt att skapa en exponentiell regressionslinje i stället för en linjär. Det är något som kan passa vissa datamodeller bättre än andra. För dessa variabler så ges intrycket att den linjära modellen bättre följer data än den transformerade. Det är något som kommer återkopplas senare i analysen.

### 5.1 – Visualisering

Nedan finns en visualisering av alla tre variabler. För att få en bild för respektive variabel enskilt se bilaga 2.



Figur 5 – Vi ser respektive variablers spridning av data samt deras linjära regression med dess konfidensintervall. Deras transformerade regressionslinje finns även visualiserad.



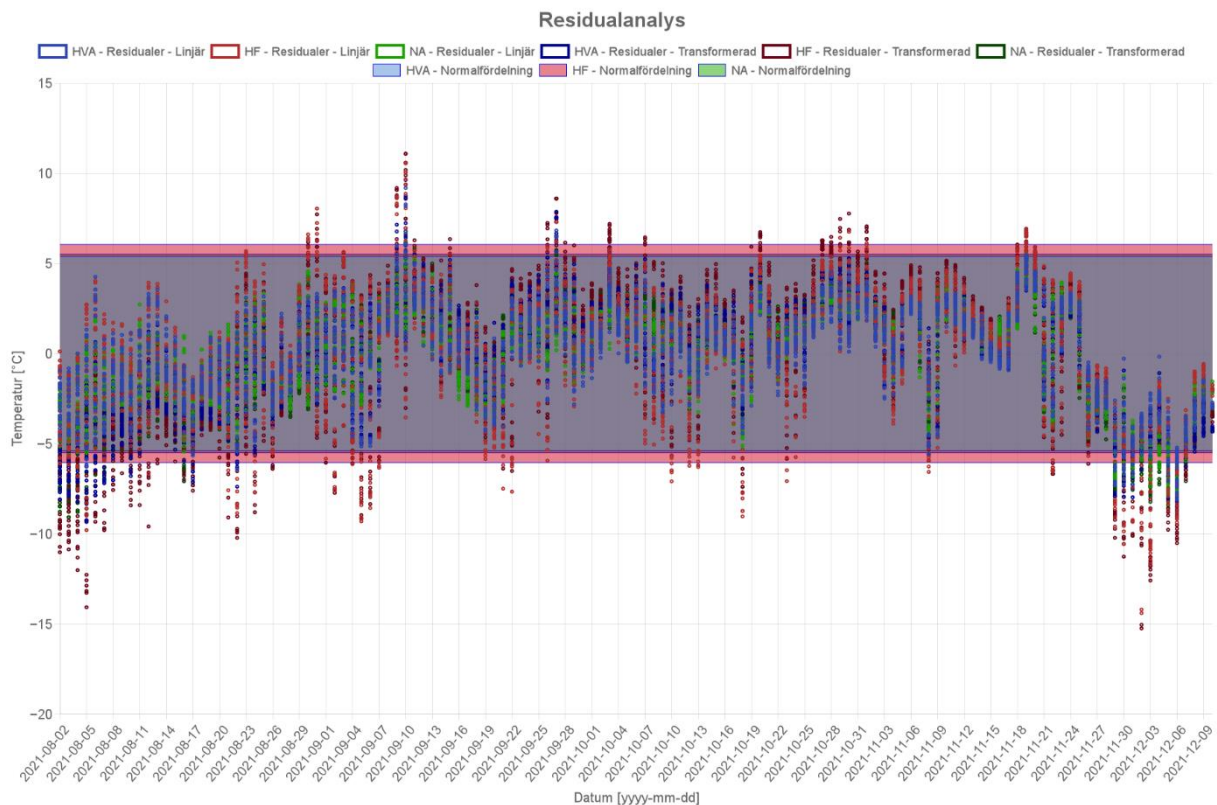
## 6 – Residualanalys

För att se hur mycket värdena för den linjära och transformerade modellen avviker från värdena som har uppmätts så utförs en residualanalys. Vi kan då enligt tabell på kapitel 6.2 se att för samtliga variabler så är variansen för data markant mindre för den linjära datamodellen. Det är även något som går att visuellt se i grafen på kapitel 6.1 samt bilaga 3. Vilket stärker mina påstående från kapitel fem angående att linjära modellerna är bättre för min data än den transformerade datamodellen.

Vi ser även att variansen förhåller sig relativt bra till normalfördelningen för både Hallands Väderö A och Nidingen A däremot sticker variansen i väg en del för Halmstad flygplats. Det tror jag har att göra med att den mätstationen har extremvärdena som både är högst och lägst. Vilket per logik borde ge störst avvikelse.

### 6.1 – Visualisering

Nedan finns en visualisering av alla tre variabler. För att få en bild för respektive variabel enskilt se bilaga 3.



Figur 6 – En visualisering av samtliga variablers residualer både linjärt och med transformerade data.

### 6.2 - Varians

	Linjär data	Transformerad data
Hallands Väderö A	6.36	10.28
Halmstads flygplats	11.89	17.76
Nidingen A	6.61	11.02

Tabell 4 – Visar respektive väderstations varianser, för både den linjära och den transformerade modellen.

## 7 – Sammanfattning

Analysen som har gjorts har visat på att linjära modellen är den som fungerar bäst för temperaturernas förändring över tid. Vilket är logiskt då temperaturer sällan ökar eller sjunker exponentiellt utan stiger eller sjunker med relativt jämn hastighet. I *tabell 3* kan vi läsa att den sjunker med cirka  $0,13^{\circ}\text{C}$  för samtliga stationer per dag och att alla startade på en temperatur mellan  $19$  och  $20^{\circ}\text{C}$ .

Vi har även påvisat att det finns starka korrelationer mellan mätstationerna, vilket går att utläsa ur *figur 2*. Det har enligt min uppfattning att göra med att temperaturerna ofta påverkas åt samma håll, dvs sjunker temperaturen på ett ställe brukar den i regel sjunka på andra ställen som inom Sverige med. Det sambandet stärks ytterligare av att mätstationerna geografiskt är nära varandra. Med en spridning med ett avstånd på  $11$  mil mellan Nidingen A och Hallands Väderö A och Halmstad flygplats placerad där emellan.

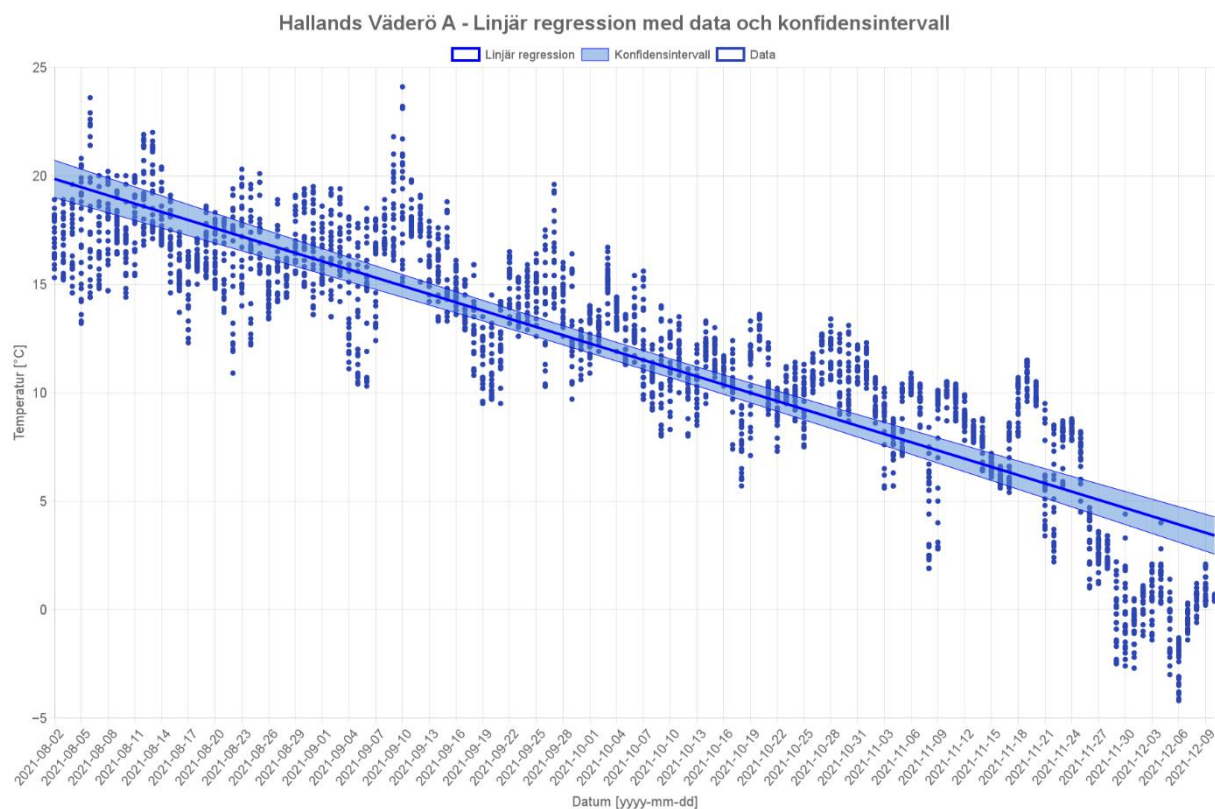
Det sista analysen påvisar är att det finns ett tydligt samband mellan mätstationernas medelvärden vilka återfinns i *tabell 2* och vart den största frekvensen av värden återfinns i histogrammet i *figur 3*.

## Referenser

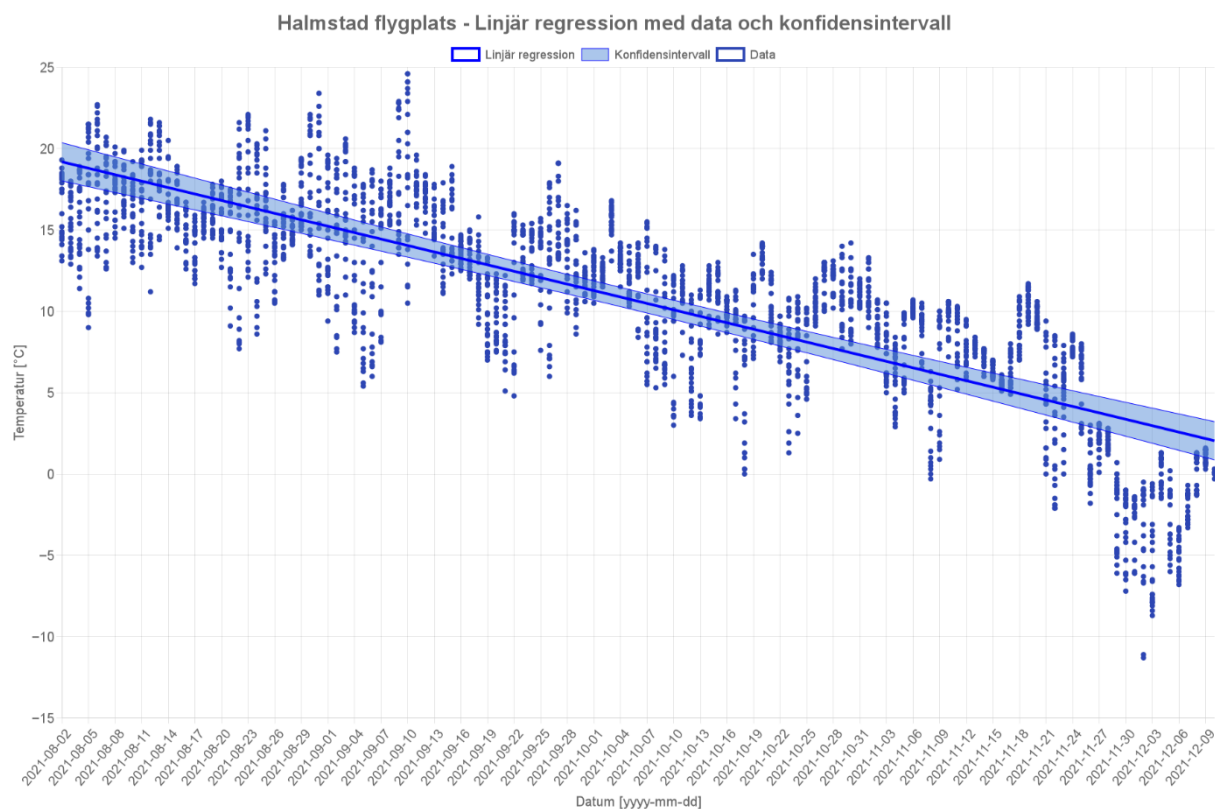
- [1] SMHI, "SMHI - Hallands väderö," [Online]. Available: <https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer#param=airtemperatureInstant,stations=all,stationid=62260>. [Använd 10 12 2021].
- [2] SMHI, "SMHI - Halmstad flygplats," [Online]. Available: <https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer#param=airtemperatureInstant,stations=all,stationid=62410>. [Använd 10 12 2021].
- [3] SMHI, "SMHI -Nidingen A," [Online]. Available: <https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer#param=airtemperatureInstant,stations=all,stationid=71190>. [Använd 10 12 2021].

## Appendix 1 – Linjär regression

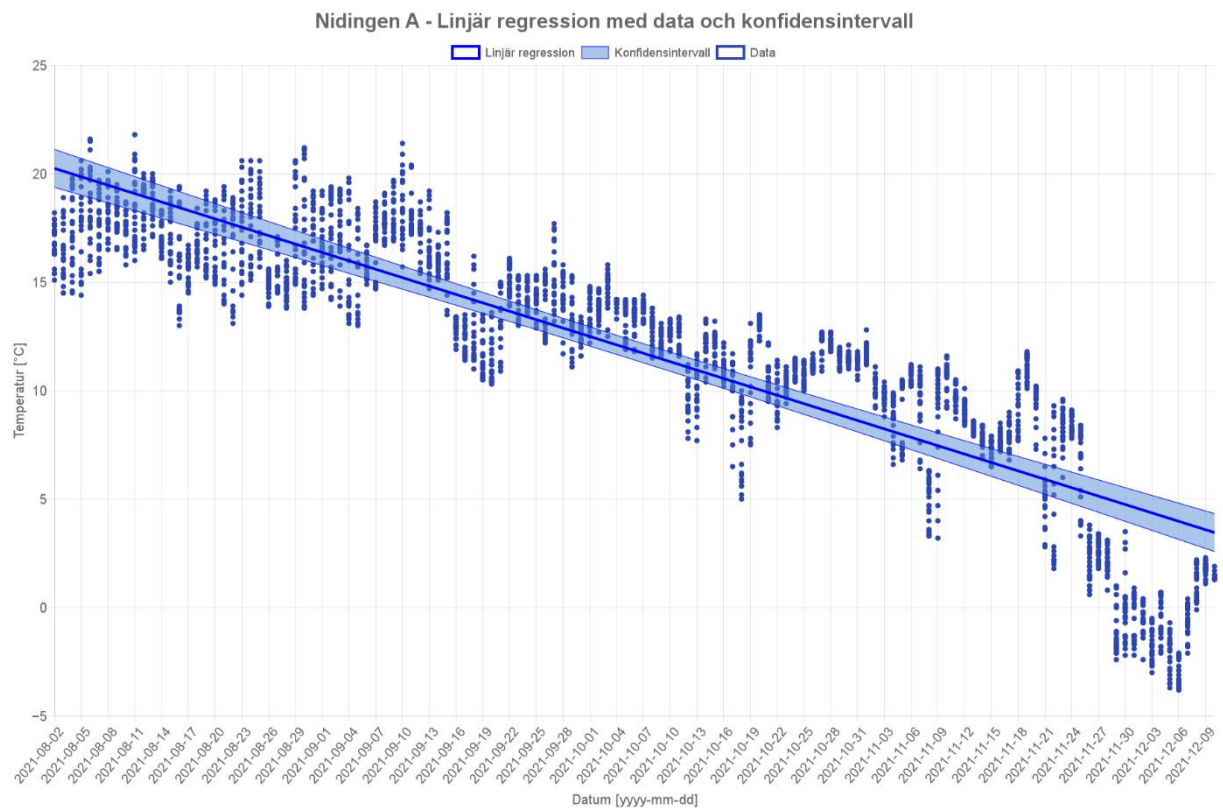
### Appendix 1.1 — Hallands Väderö A



### Appendix 1.2 – Halmstad flygplats

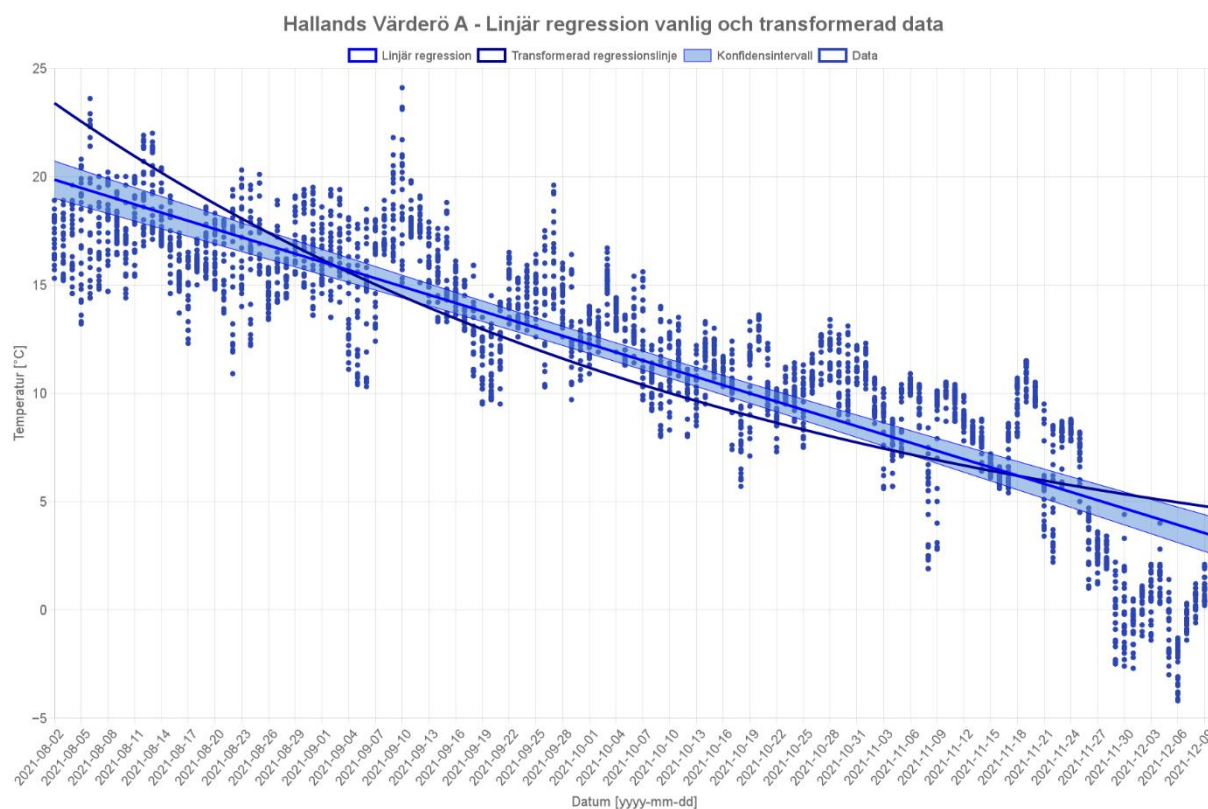


## Appendix 1.3 – Nidingen A

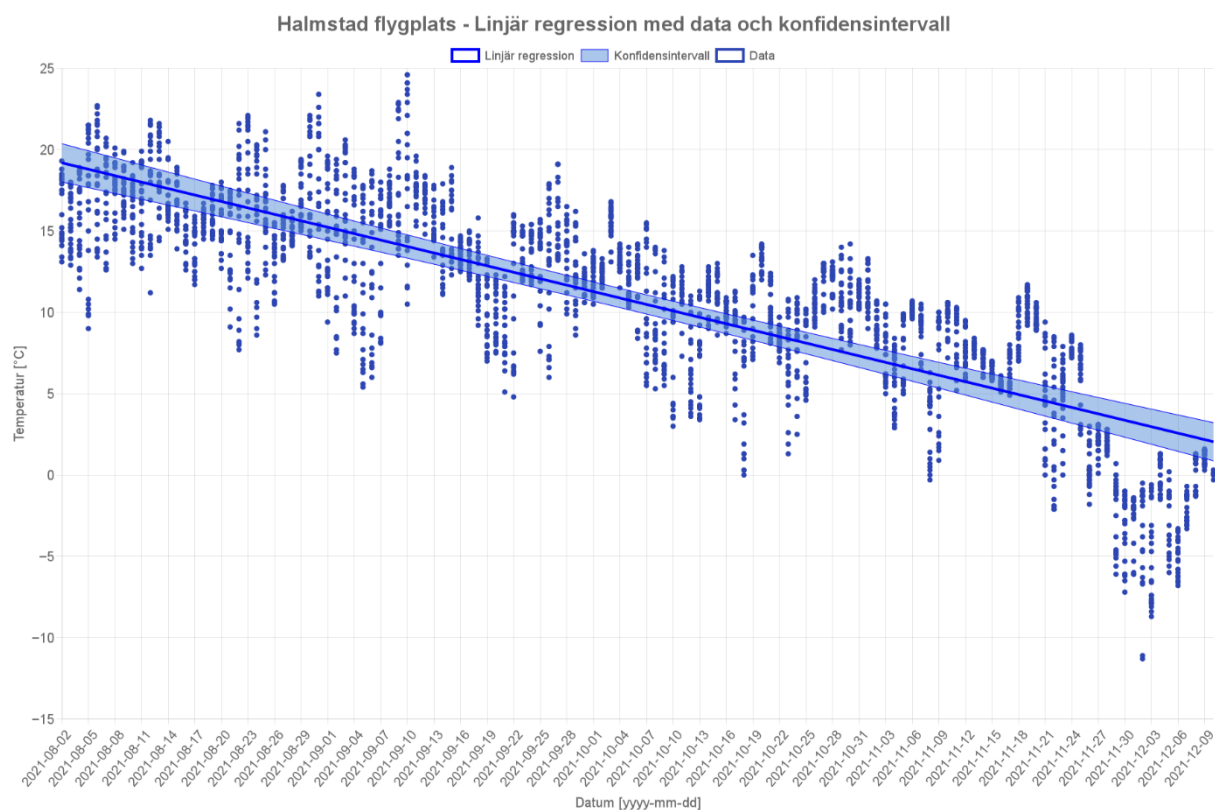


## Appendix 2 – Transformerad data

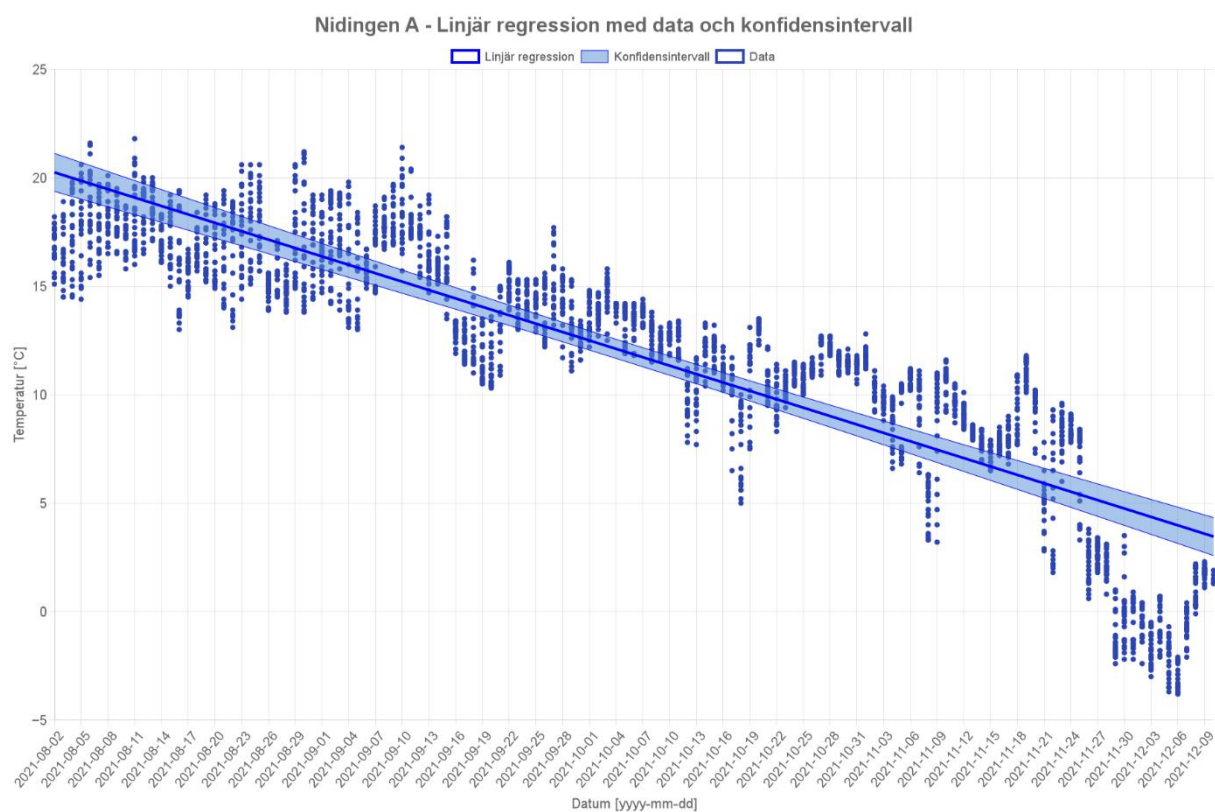
### 2.1 – Hallands Väderö A



### 2.2 – Halmstad flygplats



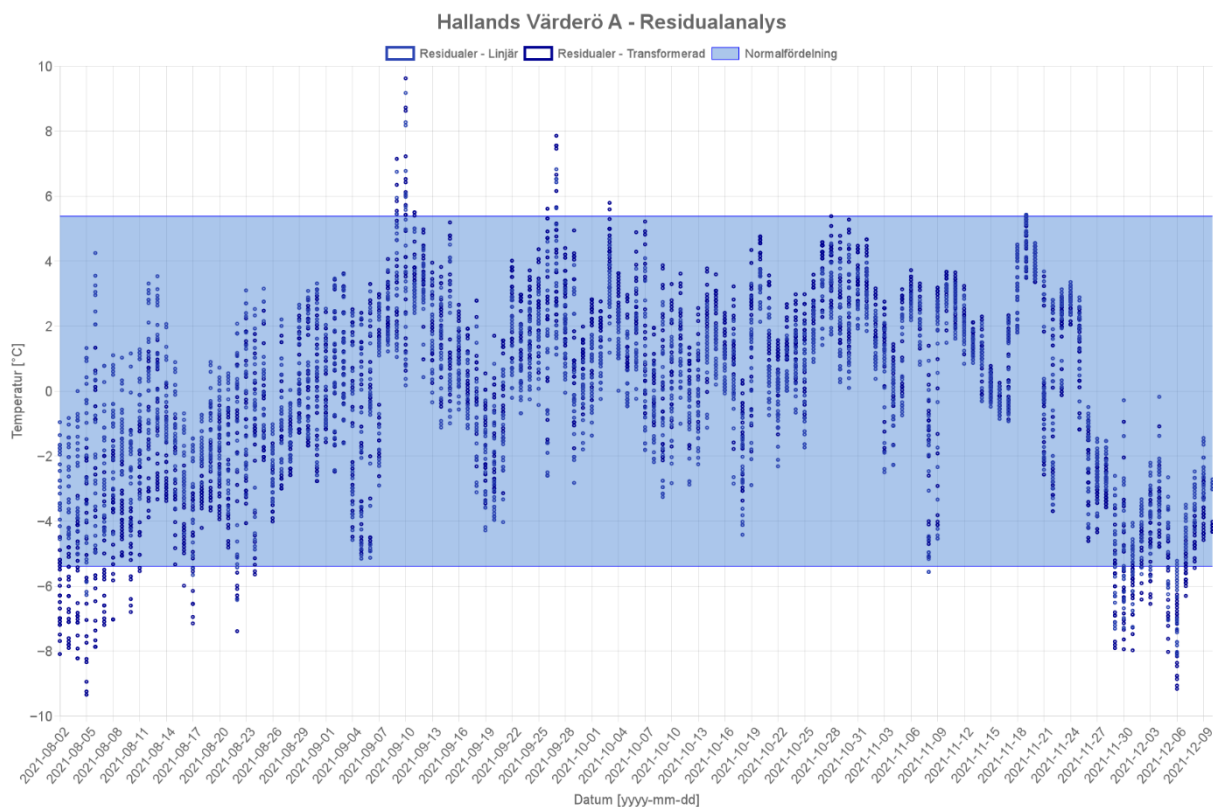
## 2.3 – Nidingen A



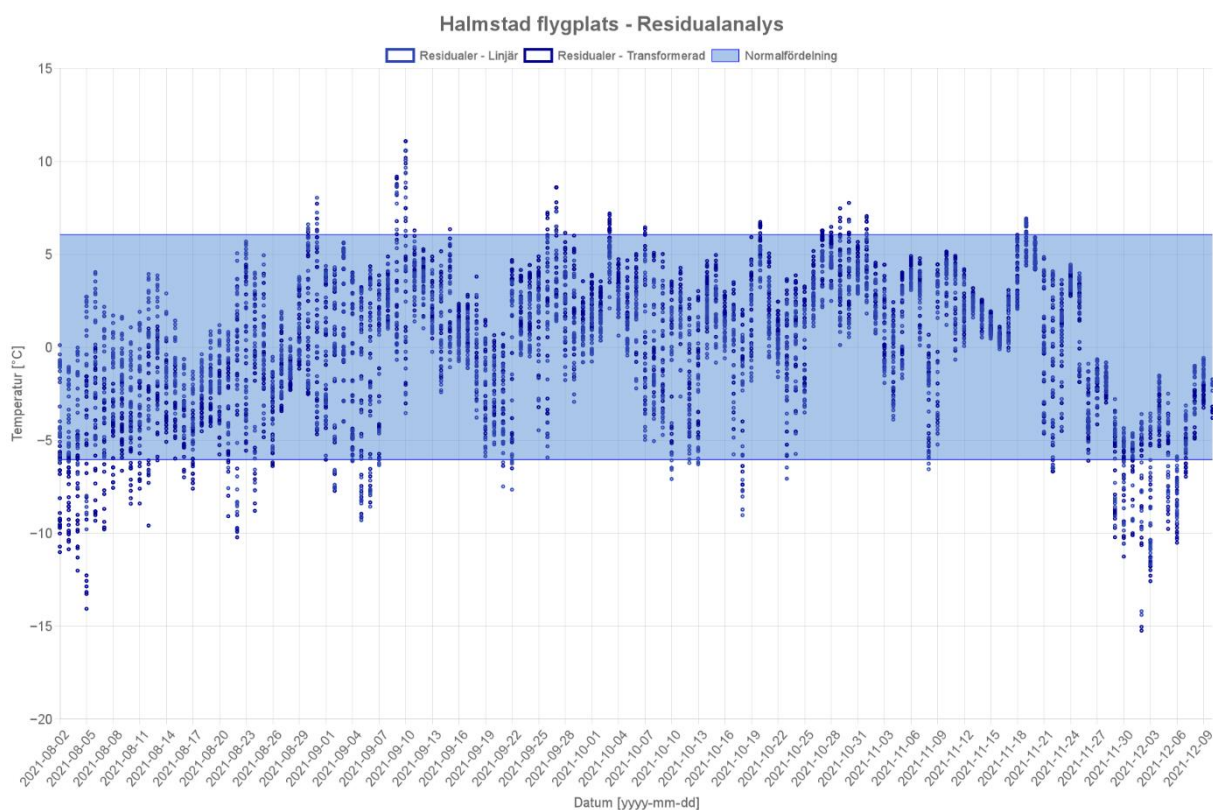


## Appendix 3 – Residualanalys

### 3.1 – Hallands Väderö A



### 3.2 – Halmstad flygplats





### 3.3 – Nidingen A

