

# Udfyldning af huller i tidsserier og udglatning af Kamstrup time aflæsninger

Dette dokument beskriver processen for udfyldning af huller i Kamstrup time serier, samt udglatningen af aflæsnings værdier til nærmeste hele time.

En generel forudsætning for dette dokument er at der i alle typer delta processeringer kun ses på aflæsninger der er mere end 4 timer gamle. Dette gøres for at begrænse antallet af potentielle huller i tidsserierne, samt begrænse antallet af opdateringer i datasættet. Denne begrænsning er laves på baggrund af feltet Aflæsningstidspunkt.

## Udfyldning af huller i tidsserier

Formålet med udfyldning af huller i tidsserierne, er at sikre at der eksisterer en aflæsning for hver time for en given måler. Hvis der identificeres en manglende aflæsning, skal denne aflæsning beregnes og indsættes, baseret på værdierne af aflæsninger omkring hullet.

Alle beregnede rækker skal markeres som værende beregnet, således at senere kan identificeres og udskiftes, i det tilfælde at den manglende aflæsning bliver modtaget fra Kamstrup.

Hvis der eksisterer flere på hinanden følgende huller i serien, beregnes den første manglende aflæsning, hvorefter den næste beregnes. Dette sikrer at der altid eksisterer en aflæsning før og efter et hul.

Hvis vi tager udgangspunkt i følgende datasæt, hvor der mangler en aflæsning for timen 2018-10-08 13:00, og hvor der både er repræsenteret Tællerstande (Energi 1 Varmeenergi, Volumen 1), Fortløbende (Timetæller) og Punktaflæsninger (Temperatur 1) i værdierne.

Tid	Måler serienummer	Aflæsningstidspunkt	Energi 1 Varmeenergi	Volumen 1	Timetæller	Temperatur 1
T <sub>0</sub>	71374198	2018-10-08 11:01:00.000	252	6,08	2261	68,83
T <sub>1</sub>	71374198	2018-10-08 12:01:00.000	254	6,12	2262	69,03
T <sub>2</sub>						
T <sub>3</sub>	71374198	2018-10-08 14:02:00.000	256	6,19	2264	68,03

For alle ikke-værdi felter, anvendes værdien fra den foregående aflæsning, dvs. Måler serienummer, Enhed, Enhed2, Enhed3, Enhed4, Enhed5, Enhed6, Enhed7, Enhed8, InfoKoder, Enhed9, Maks.-effekt 1, Enhed10, Målertype.

Aflæsningstidspunkt sættes til forrige times aflæsningstidspunkt + 1 time.

Hvis vi antager at hullet til  $T_2$  identificeres, så beregnes værdierne for den manglende aflæsning på følgende måde:

## Fortløbende

Værdien for den foregående aflæsning tillægges 1, og indsættes:

$$\text{Værdi}(T_2) = \text{Værdi}(T_1) + 1$$

## Punktaflæsninger

Der anvendes et gennemsnit af værdien for den foregående aflæsning og den efterfølgende aflæsning:

$$\text{Værdi}(T_2) = (\text{Værdi}(T_1) + \text{Værdi}(T_3))/2$$

## Tællerstand

Delta værdien mellem den foregående og den efterfølgende aflæsning divideres med 2 og lægges til værdien for  $T_1$ , og dette anvendes som værdien for  $T_2$ :

$$\text{Værdi}(T_2) = \text{Værdi}(T_1) + \Delta_{T_3-T_1}/2$$

## Eksempel

Anvender vi ovenstående udregninger på eksempel data sættet, får vi følgende datasæt efter indsættelse af manglende række:

Tid	Måler serienummer	Aflæsningstidspunkt	Energi 1 Varmeenergi	Volumen 1	Timetæller	Temperatur 1
$T_0$	71374198	2018-10-08 11:01:00.000	252	6,08	2261	68,83
$T_1$	71374198	2018-10-08 12:01:00.000	254	6,12	2262	69,03
$T_2$	71374198	2018-10-08 13:01:00.000	255	6,155	2263	68,53
$T_3$	71374198	2018-10-08 14:02:00.000	256	6,19	2264	68,03

## Udglatning af time værdier til nærmeste hele time

Formålet med udglatning af time værdier i aflæsninger er at afrunde aflæsninger til nærmeste hele time både med hensyn til tidspunktet, og værdierne i aflæsningen.

Det forudsættes i dette afsnit at alle delta tid værdier regnes i minutter, og  $\Delta V_{T_1-T_0}$  angiver forskellen i værdi mellem  $T_1$  og  $T_0$ , samt  $\Delta T_{T_1-T_0}$  angiver forskellen i tid mellem  $T_1$  og  $T_0$  målt i minutter.

Hvis vi tager udgangspunkt i følgende datasæt, der forudsættes at være komplet som en konsekvens af udfyldning af huller.

Tid	Afrundes til	Måler serienummer	Aflæsningstidspunkt	Energi 1 Varmeenergi	Volumen 1	Timetæller	Temperatur 1
T <sub>0</sub>	Er afrundet	71374198	2018-10-08 11:00:00.000	252	6,08	2261	68,83
T <sub>1</sub>	12:00	71374198	2018-10-08 11:43:00.000	257	6,12	2262	69,03
T <sub>2</sub>	13:00	71374198	2018-10-08 13:12:00.000	285	6,155	2263	68,53
T <sub>3</sub>	14:00	71374198	2018-10-08 13:52:00.000	296	6,19	2264	68,03

Det forudsættes at der altid findes en allerede udglattet værdi, som i denne tilfælde er T<sub>0</sub>. For at opfylde denne forudsætning afrundes den globalt set første aflæsning for en specifik måler til nærmeste hele time, uden at ændre på værdierne.

I resten af eksemplerne arbejder vi med "Energi 1 Varmeenergi" som eksempel på en værdi der skal udglattes.

Udglatning falder i 2 cases, hvis aflæsningstidspunkt er før den time den skal afrundes til, og hvis aflæsningstidspunkt er efter. De to cases behandles separat herunder.

### Aflæsning før ønsket hel time

I ovenstående datasæt skal T<sub>1</sub> afrundes til 12:00, dvs. efter afrunding skal værdierne afspejle hele tilvæksten fra 11:00 til 11:43, og dertil skal lægges 17 minutter af den efterfølgende aflæsning.

Hvis vi bruger "Energi 1 Varmeenergi" som eksempel for udglatning,

For at opnå dette skal værdierne for T<sub>2</sub> normaliseres over minutterne i aflæsningen, og da aflæsningen strækker sig over 89 minutter og tilvæksten i perioden er 28, giver det en fordeling pr. minut på  $28/89 = 0,315$ .

Den tilvækst der skal lægges til T<sub>1</sub> er derfor  $17 * 0,315 = 5,35$

Tilvæksten mellem T<sub>0</sub> og T<sub>1</sub> er 5

Tilvæksten for den afrundet T<sub>1</sub> bliver derfor:

$TilvækstAfrundet(T_1) = 5 + 5,35 = 10,35$

Da det er en tællerstand, lægges tilvæksten til tællerstanden for den foregående aflæsning, for at beregne værdien af tælleren for den afrundede T<sub>1</sub>:

$VærdiAfrundet(T_1) = Værdi(T_0) + TilvækstAfrundet(T_1) = 252 + 10,35 = 262,35$

Eller som generaliseret formel:

$$\text{VærdiAfrundet}(T_1) = \text{Værdi}(T_0) + \Delta V_{T_1-T_0} + (\Delta V_{T_2-T_1} / \Delta T_{T_2-T_1}) * (60 - \Delta T_{T_1-T_0})$$

## Aflæsning efter ønsket hel time

Hvis vi betragter følgende data sæt, hvor T1 nu er udglattet til 12:00:

Tid	Afrundes til	Måler serienummer	Aflæsningstidspunkt	Energi 1 Varmeenergi	Volumen 1	Timetæller	Temperatur 1
T <sub>0</sub>	Er afrundet	71374198	2018-10-08 11:00:00.000	252	6,08	2261	68,83
T <sub>1</sub>	12:00	71374198	2018-10-08 12:00:00.000	260,44	6,12	2262	69,03
T <sub>2</sub>	13:00	71374198	2018-10-08 13:12:00.000	285	6,155	2263	68,53
T <sub>3</sub>	14:00	71374198	2018-10-08 13:52:00.000	296	6,19	2264	68,03

Her ønsker vi at udglatte T<sub>2</sub> til 13:00, dvs. der skal fratrækkes 12 minutter fra værdien af T<sub>2</sub> relativ til tilvæksten.

Vi starter med at normalisere forbruget for T<sub>2</sub> over antallet af minutter, da antallet af minutter mellem 12:00 og 13:12 er 72, og tilvæksten i perioden er 24,56 giver det følgende:

Normaliseret pr. minut:  $24,56 / 72 = 0,34$

Den normaliserede tilvækst for T<sub>2</sub> beregnes som  $60 - \text{antal minutter T}_2 \text{ spænder over}$ , dvs.  $\Delta T_{T_2-T_1}$ :

$$\text{TilVækstAfrundet}(T_2) = (60-72)*0,34 = -4,09$$

For at beregne den udglattede værdi for T<sub>2</sub> lægges værdien for T<sub>2</sub> sammen med tilvæksten:

$$\text{VærdiAfrundet}(T_2) = \text{Værdi}(T_2) + \text{TilVækstAfrundet}(T_2) = 285 + (-4,09) = 280,91$$

Eller som generaliseret formel:

$$\text{VærdiAfrundet}(T_2) = \Delta \text{Værdi}_{T_2-T_1} + ((60 - \Delta T_{T_2-T_1}) / \Delta T_{T_2-T_1}) * \Delta V_{T_2-T_1}$$

## Definition af start og slut tid for databehandling

Dette afsnit beskriver hvordan start og slut tid for databehandling defineres i relation til tidspunktet for aflæsninger.

Det antages at databehandlingen kan opsplittes i to separate operationer, initial processering af det eksisterende data sæt, og behandling udført på ny indkomne aflæsninger (herefter benævnt som delta processering).

Det antages ligeledes at i det eksisterende (ikke-delta) datasæt er den sidst indsatte aflæsning (max aflæsningstidspunkt), en ikke beregnet aflæsning. Dette garanteres yderligere af definitionen af slut tidspunkt for initial processering, beskrevet herunder.

I det efterfølgende anvendes følgende sekvens af aflæsninger for en enkelt måler:

Kundenummer	Måler serienummer	Aflæsningstidspunkt	Energi 1 Varmeenergi
213928	71164164	2018-02-16 13:15:00.000	133
213928	71164164	2018-03-13 12:35:00.000	1725
213928	71164164	2018-03-20 14:15:00.000	2181
213928	71164164	2018-03-23 11:07:00.000	2318
213928	71164164	2018-03-29 09:38:00.000	2619
213928	71164164	2018-05-07 03:59:00.000	3962
213928	71164164	2018-05-18 09:26:00.000	4136
213928	71164164	2018-05-18 23:55:00.000	4148
213928	71164164	2018-05-19 01:26:00.000	4149
213928	71164164	2018-05-19 03:44:00.000	4151

For at kunne gennemføre identifikation af huller i tidsserien og udglatning af værdier er det kritisk at få defineret start og sluttidspunkt for operationen, for at kunne afgrænse tidsintervallet som operationen skal udføres på.

### Initial processering

Den initiale processering er defineret ved at operationen skal udføres på hele det eksisterende datasæt, dvs. alle eksisterende aflæsninger for en måler.

Hvis ikke en stringent start tid defineres vil operationen identificere timer før første aflæsning som huller i data, og dermed indsætte rækker på et tidspunkt hvor måleren ikke er taget i drift.

Tilsvarende, hvis måleren har været i test drift i en kort periode, dvs. der foreligger målinger for denne kort periode, og derefter har været uden signal i en længere perioden inden den endelige idrifttagning, så bør dette gab i aflæsninger ikke indgå i beregningen af huller i tidsserien.

Start tidspunktet for hver måler udlæses fra en tabel der skabes af Aalborg Forsyning, denne tabel kan efterfølgende bruges til registrering af den sidst behandlede aflæsning for hver måler.

Definitionen på slutdato defineres som max af aflæsningstidspunkt for den pågældende måler.

### Delta processing

Delta processing er defineret ved at køre på de ny indkomne aflæsninger. Denne processing kan kræve anvendelse af aflæsninger fra det eksisterende datasæt er det nødvendigt med en stringent definition af starttid for at afgrænse hvor stor en del af det eksisterende datasæt der skal tages i betragtning i processeringen.

Start tidspunkt for delta processing for en specifik måler defineres som aflæsningen med maksimum aflæsningstidspunkt i det eksisterende datasæt.

Slut tidspunkt defineres som max aflæsningstidspunkt i delta datasættet.

## Opdatering af beregnede aflæsninger

Grundet manglende leverance sikkerhed af aflæsninger fra Kamstrup, kan følgende situation indtræffe:

Måler serienummer	Aflæsningstidspunkt	Energi 1 Varmeenergi	Volumen 1	Timetæller	Temperatur 1	IsComputed
71374198	2018-10-08 11:00:00.000	252	6,08	2261	68,83	0
71374198	2018-10-08 12:00:00.000	260,44	6,12	2262	69,03	1
71374198	2018-10-08 13:00:00.000	285	6,155	2263	68,53	1
71374198	2018-10-08 14:00:00.000	296	6,19	2264	68,03	0

I ovenstående datasæt er aflæsningerne blevet processeret igennem både udfyldning af huller i tidsserier og udglatning af data.

IsComputed signalerer hvorvidt en aflæsning er beregnet som følge af udfyldning af huller, 0 er ikke beregnet, 1 er beregnet.

I det tilfælde at en aflæsning til tidspunkt 2018-10-08 12:00:00 kommer ind efter at aflæsningen 2018-10-08 14:00:00 er indsat i datasættet, opdateres rækken for 12:00:00 til værdierne for den nyindkomne aflæsning, og flaget IsComputed sættes til 0. Herefter genberegnes de efterfølgende rækker ikke, dette fører til en mindre degradering af datakvalitet, men dette accepteres af Aalborg Forsyning.

Denne operation skal logges i en separat tabel, således at der kan laves statistik over hvor mange målere dette sker for.