



DNMI

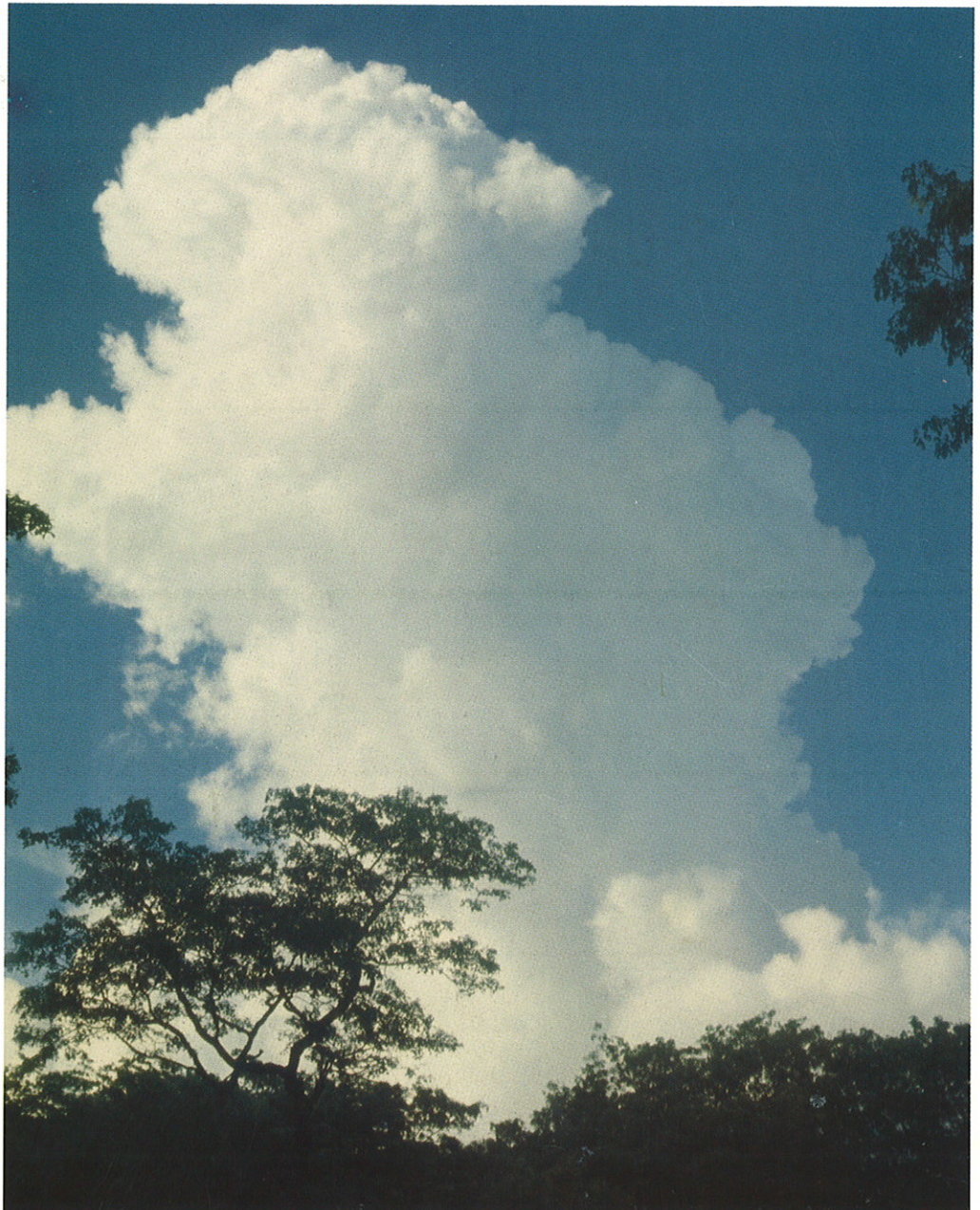
Det norske meteorologiske institutt

Report no. 17/99

KLIMMA

Interpolasjon og kvalitetskontroll i **KLIBAS**

Petter Øgland



DNMI - REPORT

ISSN 0805-9918

NORWEGIAN METEOROLOGICAL INSTITUTE
BOX 43 BLINDERN N-0313 OSLO

REPORT NO.

17/99 KLIMA

PHONE: +47 22 96 30 00

DATE

May 14 1999

TITLE

Interpolasjon og kvalitetskontroll i KLIBAS

AUTHOR

Petter Øgland

PROJECT CONTRACTOR

DNMI - Climatology Division

SUMMARY

I likhet med alle andre informasjonsbedrifter er det naturlig å forvente at store deler av det arbeidet som utføres av DNMI i dag vil bli automatisert innen de neste 25 år. Foredraget i denne rapporten beskriver hvordan utarbeidelse av data-prosessering ved Klimaavdelingen (kontroll og korreksjon av meteorologiske data) styres ved objektive metoder som bruk av statistisk prosesskontroll og kunstig intelligens for å imøtekomme denne utviklingen.

Dataprosesseringen er styrt av kvaliteten på rådata og hvilke krav som stilles til kvalitet på klimaproduktene i tillegg til hvordan selve prosesseringen utføres per se.

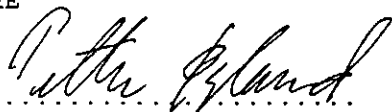
Meteorologiske data foreligger som sanntidsdata og klimadata. I arbeidet med automatisering av rutiner for bearbeidelse av sanntidsdata har man tatt i bruk feilsøkningsprogrammer og korreksjonsprogrammer (statistisk interpolasjon, bruk av modelldata og algoritmisk feilkorreksjon). Automatisering av klimadatarutinene vil basere seg på erfaring med sanntidsdata.

Klimaavdelingen deltar i et nordisk prosjekt, NORDKLIM, for samkjøring av videre kvalitetskontrollutvikling.

KEYWORDS

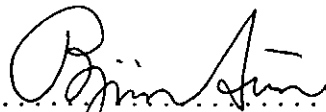
1. meteorologisk datakontroll
2. klimadatabaser
3. klimakollokvium
4. jobb-beskrivelse

SIGNATURE



Petter Øgland

Research Scientist



Bjørn Aune

Head of Climatology Division

Interpolasjon og kvalitetskontroll i KLIBAS

(Prosessering av meteorologiske data)

Klimakollokvium 18. mai 1999

1. 25 år inn i fremtiden
2. Dataprosessering
3. Datainnsamling
4. Utarbeidelse av statistikker
5. Prosessering av sanntidsdata
6. Prosessering av klimadata
7. Veien fremover

25 ÅR INN I FREMTIDEN

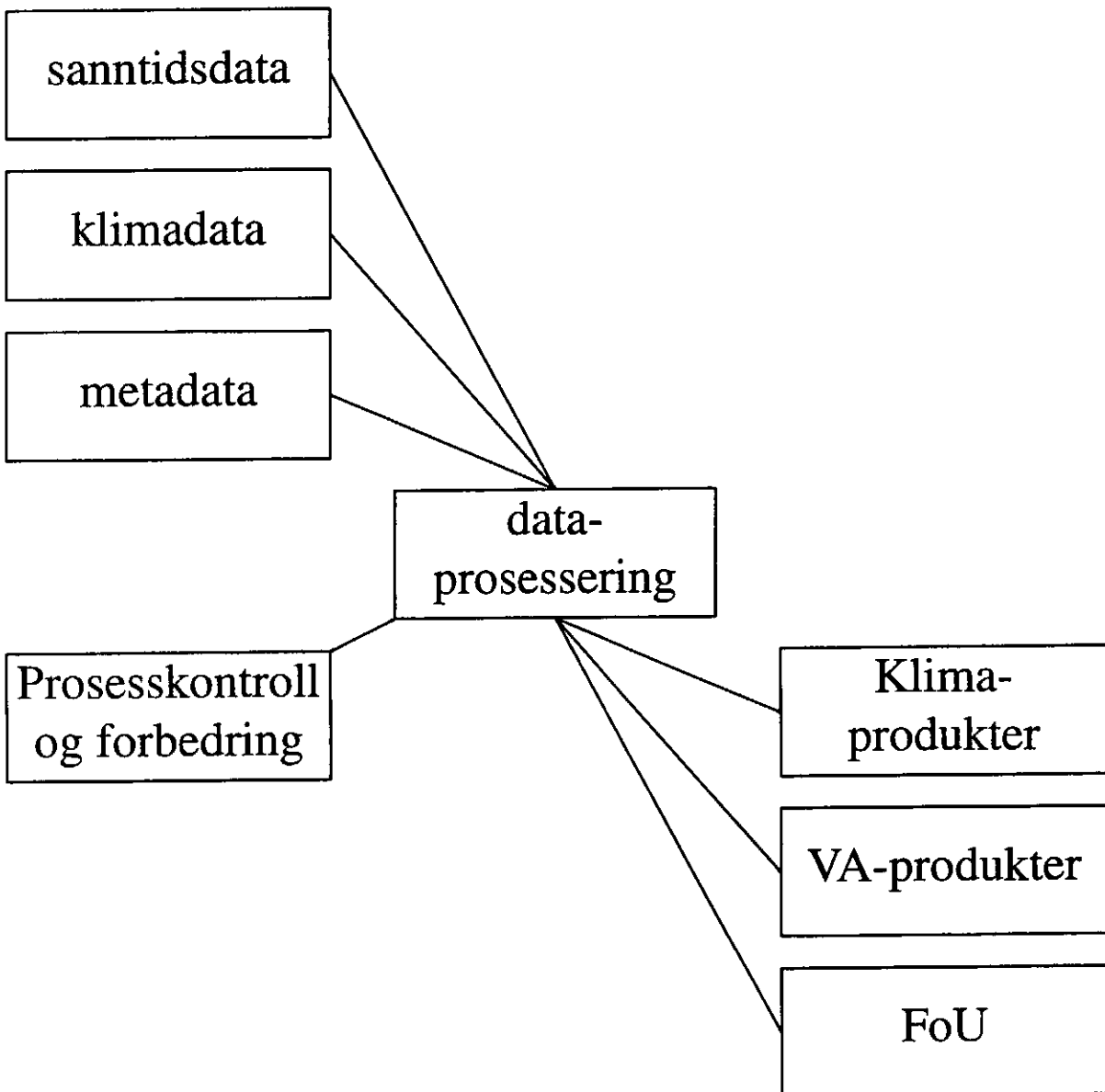
Dersom DNMI eksisterer om 25 år er et mulig scenarium som følger:

- (1) Observasjon av standard værelementer er fullstending automatisert (AWS).
- (2) Standard bearbeidelse, dvs. kvalitetskontroll og utfylling av mangler/feil, er fullstending automatisert.
- (3) Utarbeidelse av standard klimatologiske oversikter og statistikker (kjernetjenesten) er automatisert og gratis tilgjengelig via internett e.l.

Den menneskelige innsatsen vil bestå av å forbedre systemer og kartlegge hvordan vi ønsker systemene videreutviklet.

DATAPROSESSERING SOM EN DELPROSESS

Rammebetingelser for dataprosessering: (1) rådatakvalitet, (2) krav til produktkvalitet og (3) intern forskning og utvikling.



LAGRING AV SANNTIDSDATA

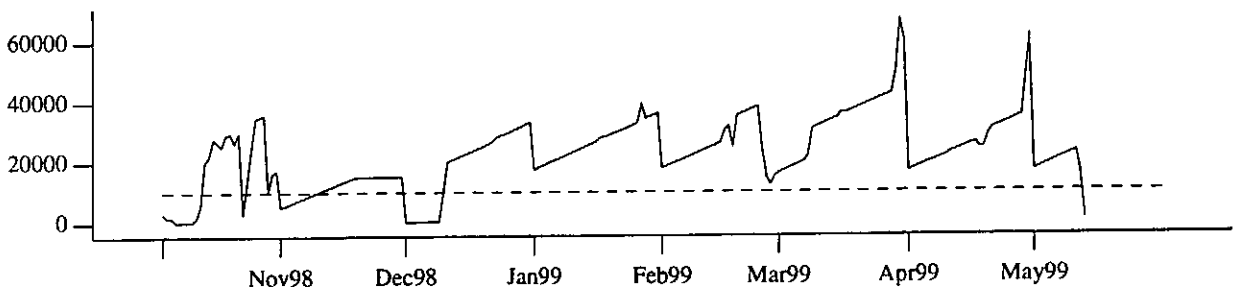
Innsamling av sanntidsdata mot KLIBAS gjøres ved automatiske rutiner som styres fra operativsystemet på databasemaskinen (crontab).

- (1) SYNO_INN. Innlasting av SYNOP-data fra binære syno-filer inn i databasetabeller SYNOP, SYNOP2, TELE og TELE2. Programmet kjøres hvert tiende minutt døgnet rundt. 209 stasjoner pr mai (158 norske).
- (2) META_INN. Innlasting av METAR-data fra binære meta-filer inn i ALF. Programmet kjøres hver tredje time. 58 stasjoner pr mai.
- (3) PIO_INN. Innlasting av PIO-data og halv-automatiske værstasjoner ("hybrider") fra månedsfiler i ascii-format inn i databasetabellen PIO. Programmet kjøres hver tredje time. 27 stasjoner pr mai (13 PIO + 14 SAWS).
- (3) AUTO_INN. Innlasting av timesdata fra automatiske værstasjoner fra månedsfiler til arbeidslager ALA og hovedlagertabeller. Programmet kjøres hver natt. 62 stasjoner pr april/mai (40 DNMI + 10 ITAS + 7 Vegvesen + 5 MDS). Kun DNMI og ITAS i historisk lager.

INNLASTINGSKONTROLL: PIO_INN

For alle de fire rutinene må det gjøres en sjekk på at dataformat som blir lest er kompatibelt med lagringsformat i databasen. Verdier som ikke lar seg lagre blir skrellet vekk.

Nedenfor vises en formatfeil-analyse som benyttes for kontroll av PIO-innlastingen.



Day by day counting of defects according to internal quality control in PIO_INN. Dashed line shows upper control limit.

Top ten most difficult stations according for format problems this month:

- 23420 FAGERNES, 41 problems.
- 69370 MERÅKER - UTSYN, 34 problems.
- 15730 BRÅTÅ - SLETTOM, 26 problems.
- 37230 TVEITSUND, 26 problems.
- 65940 SULA, 25 problems.
- 28800 LYNGDAL I NUMEDAL, 25 problems.
- 88690 HEKKINGEN FYR, 24 problems.
- 45880 FISTER - TØNNEVIK, 21 problems.
- 80610 MYKEN, 19 problems.
- 16610 FOKSTUA II, 19 problems.

Listing of format problems for 23420 FAGERNES.

- 23420: 12 cases having 15:00 not defined in ALV_PARA.
- 23420: 12 cases having 09:00 not defined in ALV_PARA.
- 23420: 11 cases having 03:00 not defined in ALV_PARA.
- 23420: 3 cases having 23:00 not defined in ALV_PARA.
- 23420: 1 case having 02:00 not defined in ALV_PARA.
- 23420: 1 case having 08:00 not defined in ALV_PARA.
- 23420: 1 case having 20:00 not defined in ALV_PARA.

VÆROVERSIKTER OG STATISTIKKER

Avhengig av hvem som benytter DNMI's produkter vil det være tilknyttet forventinger om at utskriftene er mer eller mindre korrekte. Spesielt i to tilfeller benyttes denne forventingen til å styre kvalitetskontrollen av sanntidsdata.

- (1) KA_H_STAT (klimatologisk hurtigoversikt). På samme måte som eksempelet med formatkontroll for PIO_INN, så kjøres en daglig hurtigoversikt med opptelling av antall mangler og feil i henhold til visse kriterier. Programmet lister ut og rangerer stasjoner ut fra hvor dårlig de faller ut i feilstatistikken.

- (2) STATUT (været siste 30 døgn). Samme prinsipp som for hurtigoversikten benyttes her. En automatisk kontroll leser daglig korrektur på siste-30-døgn-statistikken og rangerer stasjoner ut fra hvor dårlig de faller ut.

KVALITETSKONTROLL-SPEKIFIKASJONER

Siden 1992 har Klimaavdelingen vært involvert i spesifisering av kvalitetskontroll av værdata i flere runder.

- (1) **KLIBAS 7.2 (1992-93)**. Intern spesifisering for kvalitetskontroll av samtlige værdata, hovedvekt på nedbørobservasjoner. Avsluttet 14.05.1993.
- (2) **FREYR (1996-97)**. Nordisk spesifisering for kvalitetskontroll av nedbørobservasjoner. Avsluttet 03.02.1997. Prosjektet bygget bl.a. på erfaringer fra KLIBAS.
- (3) **NORDKLIM 1.2 (1999-)**. Nordisk spesifisering for kvalitetskontroll av samtlige værdata, med hovedvekt på sanntidsdata. Startet opp 23.04.1999. Prosjektet bygger på erfaringer fra FREYR.

Spesifikasjonene har fungert veiledende i programutviklingen. I tillegg til disse tre hovedarbeidene er det utarbeidet spesifikasjoner tilknyttet kvalitetskontrollutvikling etterhver som konkrete problemer har meldt seg.

DATAPROSESSERINGSMODULER

I forbindelse med FREYR-prosjektet la man vekt på å modularisere dataprosesseringen i tre hovedkomponenter.

- (1) **FEILSØKING.** Feilaktige eller suspekterte verdier filtreres ut ved forskjellige logiske og statistiske tester, f.eks. tidskonsistens og romlig konsistens.
- (2) **KORREKSJON.** Feilaktig eller manglende verdi kan erstattes med estimert verdi i henhold til algoritme dersom en slik algoritme er konstruerbar.
- (3) **BRUKERGRENSESNIITT.** Rasjonell bruk av feilutskrifter, kart, grafer, metadata osv. til støtte i prosesseringen.

I **NORDKLIM**-prosjektet har man også valgt å modularisere dataprosesseringen langs tidsaksen:

Automatisk sanntidsdatakontroll (AQC1), automatisk klimadatakontroll (AQC2) og manuell datakontroll (HQC).

FEILSØKING I SANNTIDSDATA

Det er utarbeidet forskjellige grad av automatiske feilsøkingmetoder i de fire sanntidsdatarutinene.

- (1) TELE-RUTINEN: Formatkontroll, logisk konsistenssjekk, sprangsjekk og grenseverdisjekk for utvalgte elementer, sluttproduktkontroll.
- (2) AWS-RUTINEN: Intern logisk konsistenssjekk, sprangsjekk, grenseverdisjekk. Sjekk m.h.p. repeterende verdier. Utskrift av ekstremverdier og ekstreme timessprang.
- (3) PIO-innlasting: Formatsjekk.
- (4) METAR-innlasting: Formatsjekk.

For AWS-rutinen utarbeides det daglig statistikk basert på oppdagede feil. Daglig log sendes via elektronisk post til IA, ITAS og Klima.

Deretter følger en månedrapport som danner utgangspunkt for kvalitetssikringsmøte mellom IA og Klima.

KORREKSJON AV SANNTIDSDATA

Automatisk korreksjon av sanntidsdata gjøres foreløpig kun i TELE-rutinen. Her benytter man tre forskjellige teknikker.

- (1) **STATISTISK INTERPOLASJON.** Romlig estimering basert på veiede midler og korreksjonsfaktor (utviklet ved DNMI). Estimering av TT, TN, TX, UU, P, P0, N, RR.
- (2) **BRUK AV MODELLEDDATA.** Data leses fra MIOPDB databasen (hirlam10), TT, TD, P, N, RR. Deretter beregnes/estimeres TN, TX, UU og P0.
- (3) **REGELBASERT KORREKSJON.** Foreløpig kun summasjonsjekk for nedbør og desimalkorreksjon for temperatur.

STATISTISK INTERPOLASJON

Interpolasjonsmetoden er en videreutvikling av Einar Borviks nedbørkontroll fra 1970-tallet.

Estimatet \tilde{x} av værelementet x benytter seg både av informasjon fra nabostasjonene både i rom og tid.

$$\tilde{x}_i = \alpha \sum_{j=1}^5 w_j y_{i,j} + \beta \quad (1)$$

hvor $y_{i,j}$ er den observerte verdien ved referansestasjon nr j relativt til teststasjonen observert ved tid i . Vektene $w_j, j=1, 2, \dots, 5$ er empirisk gitte størrelser,

$$w_j = \frac{\exp(\exp(\lambda \text{corr}(x, y_j))) - e}{\sum_{k=1}^5 [\exp(\exp(\lambda \text{corr}(x, y_k))) - e]} \quad (2)$$

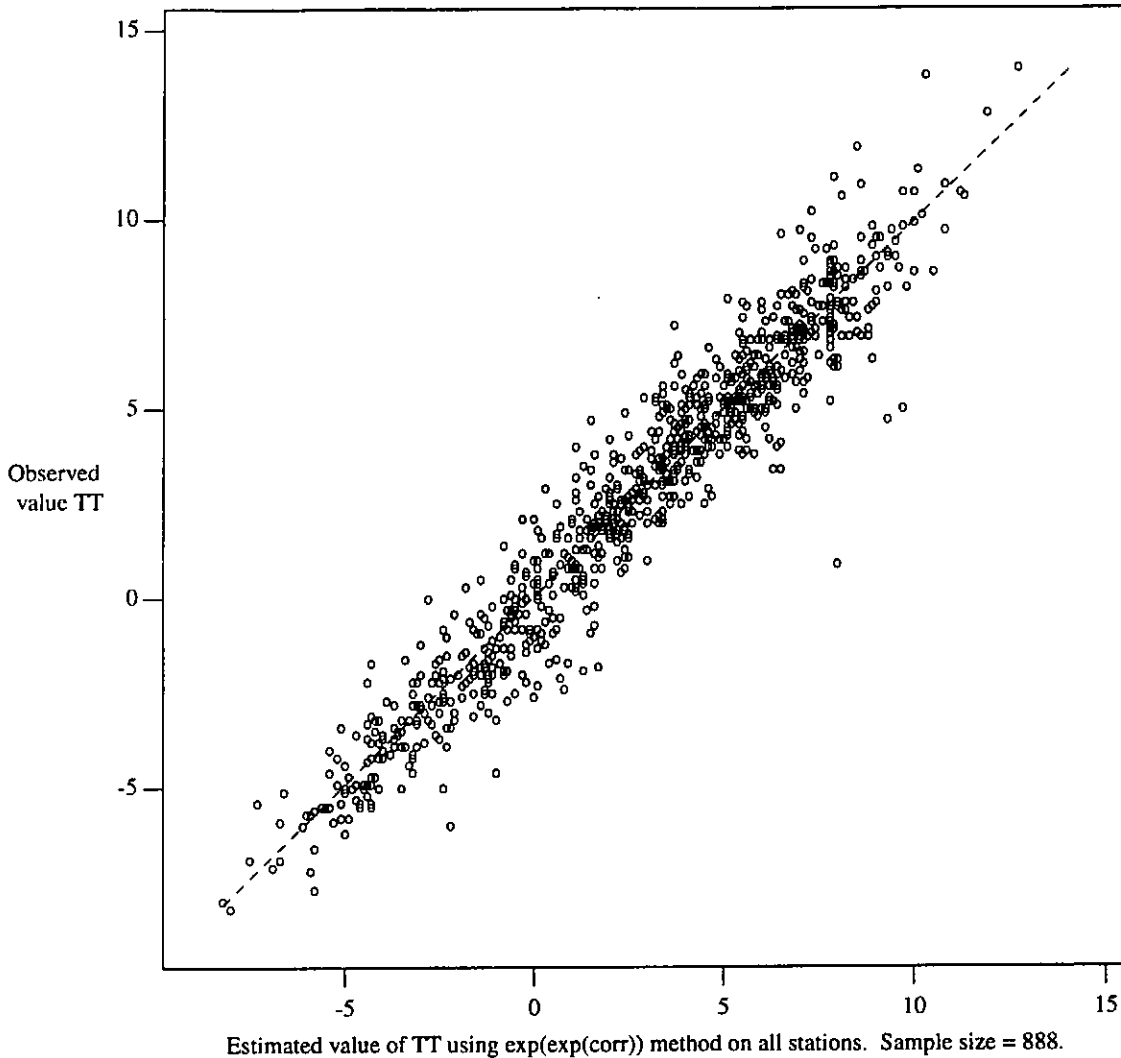
hvor

$$\lambda = \log(\log(100 + e)) \quad (3)$$

Koeffesientene α og β bestemmes empirisk. Summen av vektene er lik en.

SPREDNINGSDIAGRAM

Eksempel: Lufttemperatur TT.



| | | | | |
|-----------------------------|-----------------------------|--------------------------------|-------------------------------|---|
| Min \underline{TT} = -8.2 | Max \underline{TT} = 14.0 | Average \underline{TT} = 2.7 | Stddev \underline{TT} = 4.1 | Cov($\underline{TT}, \tilde{TT}$) = 16.3 |
| Min \tilde{TT} = -8.2 | Max \tilde{TT} = 14.0 | Average \tilde{TT} = 2.7 | Stddev \tilde{TT} = 4.2 | Corr($\underline{TT}, \tilde{TT}$) = 0.96 |

The diagonal of the plot represents the set $\underline{TT} = \tilde{TT}$, while dashed line represents the regression line

$$\underline{TT} = \alpha \tilde{TT} + \beta$$

with coefficients:

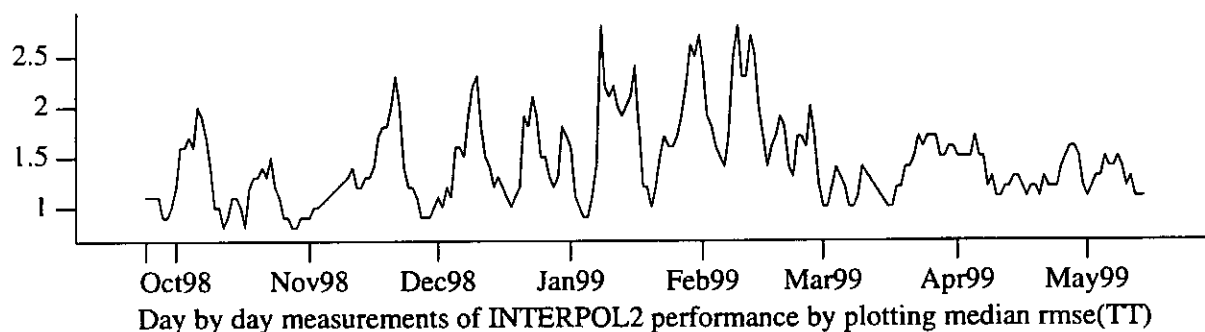
$$\alpha = 0.9930$$

$$\beta = 0.0001$$

EVALUERINGSSTATISTIKK

Metoden skal være forventningsrett (bias=0) og stde-minimerende. Feil og mangler i referansedata er et hovedproblem.

| TT at 54 stations | Sample size | alpha | beta | Bias | Stde | Rmse | Mae | Emin | Emax |
|-------------------|-------------|--------|------|------|------|------|------|------|------|
| minimum | 4 | -42.42 | -3.8 | 0.0 | 0.4 | 0.3 | 0.2 | 0.0 | 0.6 |
| median | 19 | 1.03 | 0.0 | 0.0 | 0.9 | 0.9 | 0.9 | 0.0 | 2.0 |
| maximum | 19 | 20.93 | 7.8 | 0.0 | 2.1 | 2.0 | 75.7 | 0.2 | 3.8 |
| average | 16 | 0.75 | 0.1 | 0.0 | 1.0 | 0.9 | 2.4 | 0.1 | 2.0 |
| stddev | 5 | 4.63 | 1.5 | 0.0 | 0.4 | 0.4 | 8.3 | 0.0 | 0.8 |



TN/TX Forventet RMSE = 0.6 °C.

P Forventet RMSE = 0.7 hPa.

N Forventet RMSE = 1.0 Octas.

UU Forventet RMSE = 3.7 %.

RR Forventet RMSE = 1.1 mm.

INTERPOLASJON MED MODELldata

Siste 30 døgn benyttet som grunnlagsmateriale for fortløpende evaluering.

Midlere bias(TT) = 0.2 °C, 119 stasjoner.

Midlere stde(TT) = 1.9 °C, 119 stasjoner.

Midlere bias(RR) = 0.7 mm, 112 stasjoner.

Midlere stde(RR) = 2.3 mm, 112 stasjoner.

Modelldata benyttes til å fylle inn mangler der statistisk interpolasjon kommer til kort.

PROSESSERING AV KLIMADATA

Automatisering av dataprosesseringen av klimadata følger som en konsekvens av arbeidet med sanntidsdata. Det arbeides med fire klimadatarutiner.

- (1) Værstasjoner og hybrider (Klimarutinen). Vedlikehold og videreutvikling av Lori Hålands datakontroll. Logging av kvalitetsflagg og antall defekter, og man forholder seg til utspill fra DATRUT-prosjektet.
- (2) Nedbørstasjoner. Arbeid med en helautomatisk parallellrutine (ALN2) er kommet i gang. Systematisk sammenlikning av automatisk kontroll og manuell kontroll.
- (3) Automatiske værstasjoner (Aanderaa). Elementer av statistisk prosesskontroll implementeres for å kvalitetssikre datalagringen.
- (4) AERO-innlasting (Vindregistreringer). Status er som for Aanderaa-rutinen.

VEIEN FREMOVER

Arbeidsoppgaver prioriteres på objektivt grunnlag, dvs. systemet selv avklarer hvilke programmer og oppgaver som skal behandles. Feil og advarsler logges fra 140 programmer, prioriteringer gjøres ut fra opptellinger.

Fremdriften styres av hvordan systemet modnes, dvs. ro og orden. Alt arbeid dokumenteres (analyse, design, kode, test, vedlikehold). Statusrapporter utarbeides på månedsbasis.

Videreutvikling av kvalitetskontroll og interpolasjon i KLIBAS er orientert mot to fagområder:

- (1) **STATISTISK PROSESSKONTROLL.** Systemet modelleres og evalueres kontinuerlig. Alle problemer logges, og objektiv analyse av problemene styrer videre modning og utvikling.
- (2) **KUNSTIG INTELLIGENS.** Problemer identifisert gjennom prosesskontrollen løses ved at man først analyserer hvordan problemene løses manuelt og deretter lager EDB-programmer som erstatter de manuelle rutinene i den grad det er mulig.

Klimaavdelingens kollokvieserie våren 1999

Kollokvieserien fortsetter! Datoene er nå klare. Kollokviebidrag og forslag til tema er velkomne.

Tid: Kl. 10.00

Sted: Store møterom, avlastningsbygget.

| Dato: | Navn: | Tittel/ tema: |
|-------|-----------------------|---|
| 20/1 | Alle | Planlagte forskningsaktiviteter på klimaavdelingen i 1999 |
| 10/2 | Lars Andresen | Oppsummering av prosjekt for vurdering og videreutvikling av stasjonsnettet |
| 24/2 | Bjørn Aune | ECSN og AGRØK prosjekt. En orientering. |
| 10/3 | Per Ove Kjensli m.fl. | Datavarehus |
| 24/3 | Rasmus Benestad | Evaluering av globale klimamodeller |
| 7/4 | Øyvind Nordli | Effektiv vindavkjøling i Arktis |
| 21/4 | Sofus Lystad | Entropi, informasjonsinnhold og tidsserier - UTSATT |
| 21/4 | Inger Hanssen-Bauer | Temperatur og nedbør i Norge de siste 100 år - dekadefleksjoner og trender ^{NEW} |
| 4/5 | Knut Harstveit | Vindsimulering ved Svelvik ved bruk av PHOENICS CFD-modell ^{UPDATED} NY DATO! |
| 18/5 | Petter Øgland | KLIBAS - interpolasjon og kvalitetskontroll NY DATO! |

Ansvarlige for kollokvieserien er Inger og Ole Einar.

Tidligere kollokvieserier: Våren 1998 | Høsten 1998

Kollokvier på andre avdelinger/divisjoner: FoU-divisjonen ^{UPDATED}

Denne siden er laget av OET, og sist oppdatert den 20.april 1999.