

Een Kalman-correctieschema voor de wegdektempertuurverwachtingen van het VAISALA-model

Albert Jacobs

Technisch rapport; TR-180

De Bilt, 1995

Postbus 201
3730 AE De Bilt
Wilhelminalaan 10
Telefoon 030-220 69 11
Telefax 030-221 04 07

UDC: 551.509.314
551.509.323
551.525
625.8
ISSN: 0169-1708
ISBN: 90-369-2092-2

© KNMI, De Bilt. Niets uit deze uitgave mag worden veeelvoudigd en/of openbaar gemaakt worden door middel van druk, fotocopie, microfilm, of op welke wijze dan ook, zonder voorafgaande schriftelijke toestemming van het KNMI.

Een Kalman-correctieschema voor de
wegdektemperatuurverwachtingen van het
VAISALA-model

door

Albert Jacobs, KNMI afdeling WM, 21 september 1995

Inhoudsopgave

1	Inleiding	3
2	Het Kalman-filter	7
2.1	Het Kalman-assimilatie schema	9
2.2	Het Kalman-correctie schema	11
3	Verificatieresultaten van het Kalman-filter	12
4	De operationele opzet van het Kalman-filter	17
	Dankwoord	22
	Referenties	23
A	Grafische visualisatie van de verificatie resultaten	24
B	Het format van de input/output van de Kalman programmatuur	35

1 Inleiding

De Operationele Dienst van het KNMI biedt de dienstkringen van Rijkswaterstaat meteorologische ondersteuning aan t.b.v. het Gladheidsmeldsysteem (GMS). Het KNMI levert hiertoe specifieke informatie m.b.t. de temperatuur van het wegdek¹ alsmede een aantal meteorologische parameters waaronder: de 2m-luchttemperatuur, de 2m- dauwpuntstemperatuur, de gemiddelde windsnelheid en richting op 10-m, het type bewolking, de gemiddelde hoeveelheid bewolking in okta, de neerslagkans in procenten, en het neerslag type. De meteorologische parameters komen uit DEBAV nadat deze door de meteoroloog m.b.v. een userinterface zijn bijgestuurd. Deze gecorrigeerde DEBAV-data wordt ook gebruikt om de wegdektemperatuurverwachtingen te maken. Dit gebeurt aan de hand van een fysisch model geleverd door de firma VAISALA. Het model dat voor de berekening van de wegdektemperatuur wordt gebruikt is het icecast model dat in 1984 door J. Thornes aan de universiteit van Birmingham is ontwikkeld (zie Thornes [1984]). Een nadere beschouwing van de fysica die als basis dient voor dit model treft U aan in het artikel van Rayer [1987].

Voor het bepalen van de wegdektemperatuurvoorspellingen, wordt gebruik gemaakt van de door het VAISALA-programma geleverde forecasts. Een dergelijke forecast is het resultaat van een op fysica gebaseerde 1-dimensionale diffusie vergelijking, waarbij de warmtegeleiding in de bodem over de forecast periode wordt berekend. Als randvoorwaarde voor het model dient een flux balans, de netto in- en uitstraling, aan het oppervlak van het wegdek. De netto energie flux aan het oppervlak wordt berekend aan de hand van een aantal meteorologische parameters uit DEBAV. De belangrijkste parameters die hier een rol spelen zijn de 2m-luchttemperatuur, de 2m-dauwpuntstemperatuur, de 10m-windsnelheid en richting, het type bewolking (laag, medium of hoog), de bewolgingsgraad in okta, de neerslagkans in procenten, en het type neerslag op de lokatie. Al deze parameters zijn forecasts en bepalen dus mede de kwaliteit van de wegdektemperatuurvoorspellingen. Op het moment dat het VAISALA-programma wordt opgestart zijn naast randvoorwaarden tevens beginvoorwaarden nodig in de vorm van een initiële temperatuurprofiel in de bodem. Hiervoor worden actuele temperaturen gemeten van het wegdek aan het oppervlak, en op een diepte van ca. 25 cm. De gemeten temperatuur aan het oppervlak is eigenlijk de temperatuur in de bodem op de plaats waar de sensor zich bevindt. In verband hiermee dienen we op te passen bij situaties met vorst in de grond. Het asfalt kan in staat zijn het wegdek nog net beneden de 0°C te houden, terwijl de meting aangeeft dat de temperatuur op lokatie van de sensor positief is (zie ook Minnaert [1975]). In deze situaties kan makkelijk bevriezing optreden terwijl je dat op basis van

¹Om precies te zijn worden geen verwachtingen voor de oppervlaktetemperatuur van het wegdek gemaakt, maar verwachtingen voor de temperatuur op de positie van de sensor waar de meting plaatsvindt. Deze plaats bevindt zich vaak een aantal mm. dieper in de bodem.

de sensormeting misschien niet zou verwachten. Het eindpunt van het 1-dimensionale profiel van het diffusie model bevindt zich op 115 cm diepte, en hier wordt een constante temperatuur verondersteld. De laatste aanname is acceptabel omdat zelfs bij een extreme kou de vorst niet tot deze diepte in de bodem kan doordringen. Het initiële verticale temperatuurprofiel in de bodem, noodzakelijk voor het opstarten van de forecast run, wordt nu bepaald door middel van een Lagrange interpolatie slag op basis van de gemeten oppervlaktetemperatuur, de dieptetemperatuur, en de constante temperatuur op het diepste punt van het profiel.

Een bekende eigenschap van een diffusie model is dat de rand- en beginvoorwaarden vrijwel nooit consistent zijn met elkaar. Dit verschijnsel veroorzaakt een discontinuïteit in het model, het bekende inschakel probleem. Deze dempt echter binnen een aantal uren uit, door het diffusie karakter van de vergelijking. De gebruiker dient hier, vooral 's middags, rekening mee te houden indien hij of zij uitspraken wil doen over de kwaliteit van de gemaakte forecasts. De kwaliteit van deze wegdektemperatuurvoorspellingen is in sterke mate afhankelijk van de wijze waarop de forecasts worden gemaakt. Hier speelt zowel de lengte van het voorspelinterval, het tijdstip waarop de forecast wordt opgestart, als de kwaliteit van de DEBAV invoer een rol. De gemaakte forecasts hebben een systematische gemiddelde afwijking, ook wel bias genoemd, en een bijbehorende standaard deviatie. Beide statistische parameters variëren met de geografische lokatie, het seizoen van het jaar, ze zijn afhankelijk van het tijdstip van de dag waarvoor de forecast is gemaakt, en de wijze waarop deze forecasts zijn gemaakt. In paragraaf (3) zal worden aangetoond dat de daar gebruikte VAISALA forecasts gedurende de avond, nacht en vroege ochtend een systematische negatieve bias hebben die varieert van $-0.5^{\circ}C$ tot $-1.5^{\circ}C$. De bijbehorende standaard deviatie is hier in de orde van $1^{\circ}C$. Gedurende de dag² [09^h, 17^h] is de kwaliteit van de VAISALA forecasts bijzonder slecht. Met name binnen de kritieke temperatuur range van $[-5^{\circ}C, +3^{\circ}C]$ loopt de bias op tot $+3^{\circ}C$ na het middag uur, met een standaard deviatie van eveneens $+3^{\circ}C$. De slechte performance van het voorspelmodel gedurende dit dagdeel wordt veroorzaakt door: (A) de snelle opwarming van het wegdek rond het middag uur, en de snelle afkoeling van het wegdek tegen het vallen van de avond, (B) de lage tijdsresolutie van 1 uur, en (C) het gebrek aan flexibiliteit waarmee de VAISALA forecasts kunnen worden gemaakt.

Het VAISALA model genereert forecasts tot maximaal 11^h de volgende dag. De evolutie stap van 11^h tot 12^h kan niet worden gemaakt omdat een forecast run automatisch ophoudt om 11^h. Dit betekent dat de forecasts uit het tijdsinterval [12^h, 17^h] afkomstig zijn uit een modelrun die om 12^h is opgestart, en waarvan de kwaliteit dus onbetrouwbaar is, omdat deze forecasts nog in de inschakelfase zitten. Tevens beperkt het gebrek aan flexibiliteit van de runs de maximale lengte van een voorspelinterval. De maximale voorspelhorizon van een forecast run is afhankelijk van het tijdstip van de dag waarop deze run is opgestart, en varieert van 24 uur voor de 12^h run, tot 1 uur voor de 10^h run.

²Alle tijden in deze rapportage worden uitgedrukt in UTC.

In het voorjaar van 1994 heeft het TNO een tweetal modellen voor het voorspellen van de wegdektemperaturen beoordeeld. Deze twee modellen, het fysische VAISALA model en een statistisch model gebaseerd op multi-lineaire regressie, zijn ontwikkeld door respectievelijk het KNMI en Meteo Consult. De vergelijkingsproef is gebaseerd op verificaties voor het winterseizoen 1993-1994, en de resultaten hiervan zijn samengevat in een TNO rapport, zie Jansen [1994]. Uit het onderzoek is gebleken dat tijdens de proefperiode het statistische model van Meteo Consult veel betere verwachtingen heeft geleverd dan het fysische model van het KNMI. Als reactie hierop is aan het fysische model een data-assimilatie module toegevoegd.

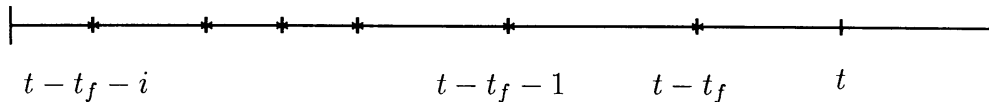
In deze rapportage wordt beschreven hoe we de VAISALA modelforecasts kunnen corrigeren op basis van een aantal historische gegevens. Er is een statistische relatie afgeleid tussen de gemaakte wegdektemperatuurvoorspellingen, en de meest recente waarnemingen van de wegdektemperatuur op de plaats van de sensor. In Persson [1990] is aangetoond dat numerieke voorspelmodellen van dit type een bias in hun temperatuur forecasts hebben die duidelijk lineair afhankelijk is van de voorspelde temperaturen zelf. Op basis van deze gegevens is het zinvol om de gemaakte forecasts te verbeteren met de correctie formule

$$\hat{T}_{fc}(t) = X_0 + X_1 \times T_{fc}(t). \quad (1.1)$$

Voor constante coëfficiënten X_0 en X_1 , is (1.1) een klassiek lineair regressie model waar de VAISALA forecasts $T_{fc}(t)$ op ieder tijdstip t worden aangepast tot de forecasts $\hat{T}_{fc}(t)$. Voor een gegeven set aan historische data proberen we de gecorrigeerde forecasts $\hat{T}_{fc}(t)$ zo goed mogelijk te fitten met de waarnemingen. Hiertoe bepalen we X_0 en X_1 m.b.v. een lineaire regressie methode. Het is altijd mogelijk om de bias over deze set volledig te elimineren, terwijl we de standaard deviatie trachten te minimaliseren. Deze aanpak leidt tot een statische regressie methode waarbij de optimale regressie coëfficiënten X_0 en X_1 vervolgens een aantal seizoenen lang gebruikt worden. Echter, normaal zullen de regressie variabelen variëren met zowel de geografische omstandigheden als de tijd van het jaar. Het afwisselend langer en korter worden van de dagen, en het afwisselen van het winter- en zomerseizoen spelen hier een rol. Het is daarom verstandig om: (1) de regressie coëfficiënten een aantal malen per jaar te verversen, of (2) de regressie coëfficiënten tijdsafhankelijk te nemen en telkens van dag tot dag aan te passen aan de zich wijzigende omgevings- en seizoensomstandigheden.

In deze rapportage introduceren we een meer dynamische methode. We beschrijven een lineaire regressie methode, waarbij de tijdsafhankelijke regressie coëfficiënten, op ieder tijdstip als er nieuwe forecasts en waarnemingen beschikbaar zijn, met behulp van een Kalman-filter worden aangepast. In paragraaf (2) zal dit Kalman-filter worden geïntroduceerd. Het filter is gebaseerd op de gemaakte wegdektemperatuurvoorspellingen $T_{fc}(t)$, en de geschiedenis van de meest recente waarnemingen voor de wegdektempera-

tuur, i.e. $\sum_{i=0}^N T_{\text{obs}}(t-t_f-i)$, met $T_{\text{obs}}(t-t_f)$, $t_f \geq 0$ de laatst beschikbare waarneming, en $N \geq 0$. Het tijdstip waarvoor de verwachting wordt gemaakt wordt hier aangegeven met t , en t_f is de lengte van het voorspelinterval, zie ook figuur (1.1). De correctie formule



Figuur 1.1: Het tijdstip t waarvoor de verwachting wordt gemaakt, en de historie $T_{\text{obs}}(t-t_f-i)$, $i = 0, 1, 2, \dots$, van de waarnemingen.

ziet er dan als volgt uit:

$$\hat{T}_{\text{fc}}(t) = X_0(t) + X_1(t) \times T_{\text{fc}}(t) + \sum_{i=0}^N X_{i+2}(t) \times T_{\text{obs}}(t-t_f-i). \quad (1.2)$$

De regressie coëfficiënten worden nu telkens dagelijks aangepast als de nieuwe waarnemingen binnen zijn. Het dagelijks automatisch bijsturen van de regressie coëfficiënten gebeurt via een recursief schema. De begin keuze voor de regressie coëfficiënten moeten we zelf maken. Deze is dus vrij onzeker, zodat het verstandig is om de gecorrigeerde forecasts pas na een periode van ongeveer 1 maand op hun kwaliteit te vertrouwen.

In paragraaf (3) worden de resultaten gegeven voor het Kalman-correctie schema toegepast op forecasts en waarnemingen uit de periode 23 november 1993 – 4 maart 1994. De resultaten worden gepresenteerd voor zowel de gehele temperatuur range, als voor de kritische range $[-5^\circ\text{C}, +3^\circ\text{C}]$. We zullen laten zien dat de bias gereduceerd kan worden tot hooguit een paar tiende graad voor voorspelintervallen tot en met 6 uur vooruit, m.u.v. forecasts die gemaakt zijn voor het dagdeel $[09^{\text{h}}, 17^{\text{h}}]$, terwijl de standaarddeviatie in de orde van $0.6^\circ\text{C} - 0.8^\circ\text{C}$ ligt, en oploopt tot 1.4°C rond het middag uur.

In de laatste paragraaf (4) wordt een vast forecast klassificatie schema en een operationeel Kalman-filter gebaseerd op deze klassificatie beschreven. Het forecast klassificatie schema specificeert een automatische produktie van forecasts. Het Kalman-filter bestaat uit een Kalman-assimilatie module waar de regressie coëfficiënten worden gecorrigeerd aan de hand van nieuwe waarnemingen, en een Kalman-correctie module waar de wegdektemperatuurvoorspellingen $T_{\text{fc}}(t)$ worden gecorrigeerd aan de hand van de lineaire regressie vergelijking (1.2) en de meest recente regressie coëfficiënten.

2 Het Kalman–filter

De essentie van het Kalman-filter is dat behalve lineaire regressie, het statische gedeelte van het filter, een dynamische evolutie vergelijking een rol speelt, die toelaat dat de regressie coëfficiënten veranderen in de tijd, en zich zo kunnen aanpassen indien de omgevings- en/of seizoensomstandigheden dat nodig achten.

De klassieke lineaire regressie vergelijking legt een lineair verband tussen een aantal predictoren $b_0(t), b_1(t), \dots, b_N(t)$, en de predictant waarde ofwel de waargenomen wegdektemperatuur $T_{\text{obs}}(t)$. We nemen aan dat de waargenomen wegdektemperatuur een stochastische variabele is die ruis bevat. We kunnen dan schrijven

$$T_{\text{obs}}(t) = \mathbf{B}(t) \times X(t) + v(t), \quad (2.1)$$

met $\mathbf{B}(t) = (b_0(t), b_1(t), \dots, b_N(t))$ de predictor waarden, $X(t) = (x_0(t), x_1(t), \dots, x_N(t))^T$ de vector met de tijdsafhankelijke regressie coëfficiënten, en $v(t)$ de ruis. We veronderstellen dat het ruisproces in (2.1) wit en Gaussisch is, met gemiddelde nul en covariantie $E[v^2(t)] = r(t) \geq 0$, een scalaire functie. $\sqrt{r(t)}$ is de standaarddeviatie voor de onzekerheid in de metingen, en de onzekerheid in de keuze van het model. Het stochastische model (2.1) wordt ook wel de observatie vergelijking genoemd. Als we nu het gemiddelde van (2.1) nemen dan vinden we ons lineaire regressie model uit de inleiding terug voor de predictor keuze

$$\mathbf{B}(t) = (1, T_{\text{fc}}(t), T_{\text{obs}}(t - t_f), T_{\text{obs}}(t - t_f - 1), \dots, T_{\text{obs}}(t - t_f - N)), \quad (2.2)$$

met $t_f \geq 0$, $N \geq 0$.

In de observatie vergelijking wordt een relatie gelegd tussen de waarnemingen en de forecasts. Tevens wordt hier ruis geïntroduceerd om enerzijds de fout in de metingen te specificeren, en anderzijds de onzekerheid in de keuze van het lineaire regressie model. De observatie vergelijking beschrijft slechts het statische gedeelte van het Kalman-filter, en is in feite niets meer dan een gewone klassieke lineaire regressie. Via een recursieve kleinste kwadraten algoritme (zie Kok en van Vliet [1994]) kunnen de optimale regressie coëfficiënten snel en efficiënt berekend worden, door de forecasts zo goed mogelijk te fitten met de waarnemingen. Hiertoe dient de gebruiker de beginkeuze van de regressie coëfficiënten, i.e. $X(0)$, in te voeren in de vorm van een gemiddelde waarde \hat{X}_0 , en een covariantie matrix P_0 . Door het invoeren van de waargenomen wegdektemperaturen worden de regressie coëfficiënten dan van dag tot dag automatisch bijgestuurd (zie ook Bryson en Ho [1975]). De keuze van de diagonaalelementen in P_0 is hierbij van uiterst belang, omdat deze aangeeft in welke mate de regressie coëfficiënten in elke iteratie stap gewijzigd

mogen worden. Hoe kleiner de diagonaalelementen, hoe minder variatie wordt toegelaten. Omdat de keuze van de beginwaarde \hat{X}_0 over het algemeen nogal onzeker is, doet de gebruiker er verstandig aan om de diagonaal elementen groot te kiezen. Deze zullen dan in de eerstvolgende iteratie slagen vrij snel afnemen en de regressie coëfficiënten zullen snel tot convergentie overgaan. De covariantie matrix P van de regressie coëfficiënten, zie voor de definitie hiervan paragraaf 2.1, wordt steeds kleiner naarmate het iteratie proces vordert, en laat dus steeds minder vrijheid toe in de variatie van $X(t)$. De covariantie matrix P is dus een maat voor de nauwkeurigheid waarmee een optimale schatter voor $X(t)$ is berekend. De uiteindelijk berekende regressie coëfficiënten zijn bij benadering de optimale, welke men op basis van een kleinste kwadraten methode voor dezelfde dataset zou hebben gevonden. Bij het gebruik van een klein aantal predictoren en redelijke aannamen voor de beginwaarden, zullen de iteratief berekende regressie coëfficiënten na ongeveer 1 maand niet meer van de optimale te onderscheiden zijn.

Een Kalman-filter voegt aan dit statische model nog een tweede, een stochastisch dynamisch model toe. Het is een vergelijking die de tijdsevolutie van de regressie coëfficiënten beschrijft:

$$X(t+1) = \mathbf{A}(t)X(t) + W(t+1), \quad (2.3)$$

met $\mathbf{A}(t)$ de systeem matrix en $W(t)$ de ruis die wit en Gaussisch is met gemiddelde nul en covariantie

$$E[W(t+1)W^T(t+1)] = Q(t+1). \quad (2.4)$$

We veronderstellen hier, dat de beginkeuze voor de regressie coëfficiënten, i.e. $X(0)$, gemiddelde \hat{X}_0 heeft en covariantie M_0 . Tevens nemen we aan dat de begintoestand $X(0)$ niet gecorreleerd is met de modelfout W , en zijn ook de fouten v en W niet onderling gecorreleerd.

Als we nu het gemiddelde en de covariantie van (2.3) nemen, dan vinden we

$$\begin{aligned} \hat{X}(t+1) &= \mathbf{A}(t)\hat{X}(t), \\ M(t+1) &= \mathbf{A}(t)M(t)\mathbf{A}^T(t) + Q(t+1). \end{aligned} \quad (2.5)$$

In tegenstelling tot de statische observatie vergelijking bepaalt de evolutie vergelijking de variatie van de regressie variabelen X in de loop van de tijd. De seizoensafhankelijkheid van de regressie variabelen wordt beschreven d.m.v. het evolutiemodel (2.3), en is daarin ook in belangrijke mate afhankelijk van de transitie matrix $\mathbf{A}(t)$. Echter, het ontbreekt ons aan kennis, zelfs in de vorm van elementaire empirische formules, om een redelijke

beschrijving te geven van de matrix $\mathbf{A}(t)$ als functie van de tijd. In al onze toepassingen nemen we dan ook $\mathbf{A}(t) = I$, de identiteit, zodat alle veranderingen in de regressie coëfficiënten worden veroorzaakt door het ruisproces $W(t)$, en dus door de covariantie vergelijking in (2.5). De gemiddelde tijdsevolutie van de regressie coëfficiënten in (2.5) is nu immers stationair geworden.

De covariantie matrix Q in de evolutie vergelijking (2.5) bepaalt hoe flexibel het filter is. Als de covarianties in Q klein zijn, dan is het filter weinig ontvankelijk voor nieuwe data en zal vasthouden aan datgene wat hij al geleerd heeft. Het filter heeft in dit geval een bijzonder lang geheugen en zal zich de gehele voorgeschiedenis van de waarnemingen en de gemaakte forecasts blijven herinneren. Het filter is stijf en reageert uiterst traag op veranderende omstandigheden in de vorm van nieuwe data. Dit probleem doet zich voor bij lineaire regressie als na verloop van tijd de covarianties P van de regressie coëfficiënten klein worden en het filter zich slechts uiterst traag aanpast aan nieuwe data met een andere systematische fout.

De evolutievergelijking van het Kalman-filter biedt ons de mogelijkheid om d.m.v. een externe actie, op ieder gewenst tijdstip in te grijpen in het systeem. We kunnen de covariantie matrix $Q(t)$ aanpassen wanneer de omstandigheden daar aanleiding toe geven. Door deze interactieve ingreep kunnen de covarianties in $Q(t)$ tijdig iets verhoogd worden, om zodoende de regressie coëfficiënten meer vrijheid te geven om zich snel te kunnen aanpassen aan de nieuwe situatie. Een voorbeeld van een dergelijk interactief systeem treft U aan in Persson [1990]. Een Kalman-filter heeft dus duidelijk een veel korter geheugen omdat het aanpassingsvermogen kan worden gecoördineerd d.m.v. bijsturing van de covarianties. Het tijdelijk vergroten van $Q(t)$ betekent dat het filter ververscht wordt, en weer jong en ontvankelijk voor nieuwe informatie is.

2.1 Het Kalman–assimilatie schema

In deze rapportage koppelen we de klassieke lineaire regressie formule (2.1) aan de evolutievergelijking (2.3). We bepalen de gecorrigeerde regressie coëfficiënten \hat{X} , op basis van een Kalman data-assimilatie schema. Daartoe assimileren we de waarnemingen $T_{\text{obs}}(t)$ in het evolutie model (2.3). Voor de uitwerking van de technische details verwijzen we naar Maybeck [1979].

Het evolutie model (2.5) is de eerste stap in ons Kalman-assimilatie schema. We schrijven deze evolutie stap als

$$\begin{aligned}\hat{X}(t+1|t) &= \hat{X}(t|t), & \hat{X}(0|0) &= \hat{X}_0, \\ P(t+1|t) &= P(t|t) + Q(t+1), & P(0|0) &= P_0.\end{aligned}\tag{2.6}$$

Hier is $\hat{X}(t+1|t)$ het voorwaardelijke gemiddelde van $X(t+1)$, onder de voorwaarde dat alle waarnemingen tot en met het tijdstip t zijn geassimileerd. Door de koppeling van het statische en dynamische model hebben we de covarianties vervangen door P . In vergelijking met het voorafgaande is $P_0 = M_0$, $P(t+1|t)$ de covariantie die hoort bij $\hat{X}(t+1|t)$, en $P(t|t)$ de covariantie van $\hat{X}(t|t)$, de gecorrigeerde $X(t)$ nadat ook de laatste waarneming $T_{\text{obs}}(t)$ is geassimileerd.

De tweede stap, de assimilatie of filter stap, is gebaseerd op het lineaire regressie model (2.1). Deze stap berekend de correcties van de regressie coëfficiënten

$$\hat{X}(t+1|t+1) = \hat{X}(t+1|t) + K(t+1) \left[T_{\text{obs}}(t+1) - B(t+1) \times \hat{X}(t+1|t) \right]. \quad (2.7)$$

In vergelijking (2.7) is $K(t+1)$ een gewichten matrix, de Kalman gain, die aangeeft hoe de regressie coëfficiënten $\hat{X}(t+1|t)$ gecorrigeerd dienen te worden m.b.v. de laatste waarneming $T_{\text{obs}}(t+1)$. Deze Kalman gain bepaalt de kwaliteit van het filter. Voor het bepalen van de Kalman gain berekenen we een optimale schatter voor de regressie coëfficiënten $\hat{X}(t+1|t+1)$. Deze schatter is de maximum likelihood schatter van de voorwaardelijke kansdichtheidsfunctie $p(X|T_{\text{obs}})$, gebaseerd op formule (2.1). Omdat alle ruisprocessen wit en Gaussisch zijn, en omdat ons model lineair is, is ook de toestandsvector X in (2.1) Gaussisch verdeeld, en wordt dus slechts bepaald door zijn gemiddelde en covariantie. De voorwaardelijke kansdichtheid is dus Gaussisch en bevat verder alle informatie die nodig is om een optimale schatter voor X te bepalen. Beschouw nu de kwadratische functie

$$\begin{aligned} J[X] = & \frac{1}{2} \left\{ \left[X - \hat{X}(t+1|t) \right]^T P^{-1}(t+1|t) \left[X - \hat{X}(t+1|t) \right] + \right. \\ & \left[T_{\text{obs}}(t+1) - B(t+1) \times \hat{X}(t+1|t) \right]^T r^{-1}(t+1) \times \\ & \left. \left[T_{\text{obs}}(t+1) - B(t+1) \times \hat{X}(t+1|t) \right] \right\}. \end{aligned} \quad (2.8)$$

De optimale schatter voor X is nu precies gelijk aan de gewogen kleinste kwadraten schatter die de functie J in (2.8) minimaliseert (zie Maybeck [1979]). Deze schatter noteren we als $\hat{X}(t+1|t+1)$, zie ook (2.7). De Kalman gain matrix $K(t+1)$ kan uit deze minimalisatie worden afgeleid, en is

$$K(t+1) = P(t+1|t)B^T(t+1) \left[B(t+1)P(t+1|t)B^T(t+1) + r(t+1) \right]^{-1}. \quad (2.9)$$

De bijbehorende covariantie van $\hat{X}(t+1|t+1)$ is nu

$$P(t+1|t+1) = P(t+1|t) - K(t+1)B(t+1)P(t+1|t). \quad (2.10)$$

Samenvattend bestaat het Kalman-assimilatie schema nu uit de evolutie stap (2.6), de filter stap (2.7), de Kalman gain update (2.9) en de covariantie die hoort bij de gecorrigeerde regressie coëfficiënten, dat is de vergelijking (2.10).

2.2 Het Kalman–correctie schema

De correctie formule waarmee de VAISALA forecasts worden gecorrigeerd, is van de vorm

$$\hat{T}_{fc}(t) = \tilde{X}_0(t) + \tilde{X}_1(t) \times T_{fc}(t) + \tilde{X}_2(t) \times T_{obs}(t - t_f), \quad (2.11)$$

waar we alleen de laatst beschikbare waarneming, $T_{obs}(t - t_f)$ voor zekere $t_f \geq 0$, uit de voorgeschiedenis hebben meegenomen (zie ook formule (1.2) met $N = 0$). Uit experimenten is gebleken dat een langere voorgeschiedenis van de waargenomen wegdektemperaturen, i.e. $N \geq 1$, geen enkele impact heeft op de kwaliteit van de gecorrigeerde forecasts, zowel wat de statistiek betreft, als de afzonderlijke forecasts. De regressie coëfficiënten in (2.11) zijn slechts benaderingen voor de optimale schatters $\hat{X}_0(t)$, $\hat{X}_1(t)$ en $\hat{X}_2(t)$, omdat de laatste pas bekend zijn als $T_{obs}(t)$ beschikbaar is, en geassimileerd is in het regressie model. Omdat $T_{obs}(t - t_f)$ de laatst beschikbare waarneming is, kunnen we $t_a = t - t_f$ beschouwen als het actuele tijdstip. De forecasts $T_{fc}(t)$ en correcties $\hat{T}_{fc}(t)$, hebben dan een voorspelinterval van t_f uur vooruit. De aanwezigheid van $T_{obs}(t_a)$ betekent niet dat we het VAISALA model dan maar op het tijdstip t_a moeten opstarten. Probeer dit ook indien mogelijk te vermijden. De zo gemaakte forecasts komen anders uit de inschakelfase, en deze kun je maar moeilijk corrigeren. De VAISALA forecasts halen we uit een run die we minimaal 6 uur eerder hebben opgestart. Tot en met het tijdstip $t_a - 1$ worden dan de VAISALA forecasts vervangen door de waarnemingen, en op t_a gebruiken we de correctie formule (2.11) met een voorspelinterval van $t_f = 0$ uur.

We gebruiken nu de correctie formule (2.11) om vanaf het actuele tijdstip $t = t_a$, de toekomst te gaan voorspellen voor $t = t_a + t_f$, met $t_f = 1, 2, 3, \dots$. Omdat de regressie coëfficiënten voor deze tijdstippen nog niet aanwezig zijn, gebruiken we als benadering regressie coëfficiënten uit het verleden. Onze keuze van regressie coëfficiënten is gebaseerd op de dagelijkse cyclus van dag en nacht, die een belangrijke rol speelt bij het opwarmen en afkoelen van het wegdek. Deze cyclus zal bij benadering ook terug te vinden zijn in de regressie coëfficiënten. In het bijzonder geldt,

$$\hat{X}_k(t) \approx \hat{X}_k(t - 24^h), \quad k = 0, 1, 2, \quad (2.12)$$

de periode is dus ongeveer 1 dag. Let wel op dat deze periodiciteit slechts bij benadering geldt, omdat er telkens een verschuiving plaatsvindt van de dagelijkse cyclus, veroorzaakt

door het afwisselend langer en korter worden van de dagen. Dit fenomeen (de lange tijdschaal) speelt een belangrijke rol wanneer we winter en zomer seizoenen met elkaar vergelijken. Bij een vertraging van 1 dag (de korte tijdschaal), zoals in (2.12), is deze verandering nog nauwelijks merkbaar. De veranderingen op lange tijdschalen worden mede mogelijk gemaakt door een optimale instelling van de covarianties in de tijdsevolutie vergelijking van het Kalman-filter. Op de korte tijdschalen worden de regressie coëfficiënten telkens van dag tot dag via het Kalman-assimilatie schema aangepast.

De optimale regressie coëfficiënten in het Kalman-assimilatie schema worden bepaald door de structuur van de regressie formule. Dit betekent dat de regressie coëfficiënten behalve van het tijdstip van de dag ook afhankelijk zijn van de lengte van het voorspelin-terval, i.e. t_f . Als gevolg van de kenmerkende dagelijkse cyclus van 24^h , zullen we de regressie coëfficiënten telkens om de 24^h aanpassen. We berekenen dus

$$\hat{X}_k(t + n \times 24^h; t_f), \quad k = 0, 1, 2, \quad n = 0, 1, 2, \dots, \quad \text{en } t_f = 0, 1, 2, \dots \quad (2.13)$$

Dit betekent dus voor ieder uur van de dag $t = 00^h, 01^h, \dots, 23^h$, en voor ieder voorspelin-terval $t_f = 0$ uur, 1 uur, \dots , een apart Kalman-filter waarin de coëfficiënten \hat{X}_0 , \hat{X}_1 , en \hat{X}_2 worden aangepast.

3 Verificatieresultaten van het Kalman-filter

In deze paragraaf geven we de verificatieresultaten van de VAISALA verwachtingen, en de door het Kalman-correctie schema gemaakte gecorrigeerde verwachtingen. Alle resultaten hebben betrekking op de periode 23 november 1993 – 4 maart 1994. De invoer voor het VAISALA programma bestaat uit de DEBAV verwachtingen voor de lokatie Soesterberg, die de randvoorwaarden bepalen, en de waargenomen wegdek- en dieptetemperatuur voor de RW12 te Maarn, die het initiële temperatuurprofiel bepaalt. Voor de verificatie worden alleen forecasts gebruikt die buiten de inschakelfase (de eerste 6 uren van iedere modelrun) liggen, m.u.v. de forecasts die gemaakt zijn voor het dagdeel $[12^h, 17^h]$.

We bestuderen de kwaliteit van de verwachtingen aan de hand van een tweetal verificatiegrootheden: (1) de gemiddelde fout in de verwachtingen (ME), ook wel bias genoemd, over de gehele verificatie periode, en (2) de standaarddeviatie (SD) van de fout in de verwachtingen, eveneens over de gehele verificatie periode. Beide verificatiegrootheden worden gegeven voor ieder uur van de dag, en voor voorspelinintervallen waarvan de lengte varieert van $t_f = 1$ uur tot en met $t_f = 6$ uur. Dit betekent dat we maximaal 6 uur

voortuit voorspellen. De presentatie van de resultaten is zowel grafisch als in tabelvorm. De grafische voorstelling geschiedt in de vorm van tijdreeksen voor de maanden januari en februari 1994; scatterdiagrammen voor de kritische temperatuurzone voor het dagdeel $[09^h, 17^h]$, de avond $[18^h, 23^h]$, en de nacht $[00^h, 08^h]$; de gemiddelde afwijking en standaarddeviatie voor voorspelintervallen ter lengte $t_f = 1$ uur tot en met $t_f = 6$ uur (zowel het gehele temperatuur bereik, als het kritische); en voor de kritische temperatuurzone $[-5^\circ C, +3^\circ C]$, tevens de gemiddelde afwijking en standaarddeviatie als functie van de lengte t_f van het voorspelinterval, voor een aantal tijdstippen van de dag. Deze laatste resultaten worden tevens in tabelvorm gepresenteerd.

Omdat de VAISALA forecasts gemaakt zijn via een vast klassificatie schema, en we de eerste uren van een forecast run liever niet gebruiken, nemen we het voorspelinterval van deze forecasts niet in beschouwing. De gemiddelde afwijking en standaarddeviatie van de forecasts zijn dan alleen afhankelijk van het tijdstip van de dag. De lengte van het voorspelinterval, t_f , heeft alleen betrekking op de Kalman-assimilaties en correcties. Voor we overgaan tot de presentatie van de resultaten, volgt nu eerst de klassificatie van de VAISALA voorspellingen.

VAISALA forecast klassificatie

1. De uurlijkse forecasts $[12^h, 17^h]$ komen uit VAISALA runs die dagelijks om 12^h worden opgestart. De invoer voor iedere VAISALA run is de waargenomen wegdek- en dieptetemperatuur van 12^h voor het bepalen van het initiële temperatuurprofiel, en de DEBAV van 12^h voor het bepalen van de randvoorwaarden.³ De forecast run loopt tot 11^h de volgende dag. Let op, de gemaakte forecasts zitten in de inschakelfase van de runs.
2. De uurlijkse forecasts $[18^h, 23^h]$ komen uit VAISALA runs die dagelijks om 12^h worden opgestart, om zodoende het inschakeleffect van de eerste uren te vermijden. Als invoer worden de waargenomen wegdek- en dieptetemperatuur van 12^h gebruikt, de DEBAV van 12^h voor de randvoorwaarden op het interval $[12^h, 17^h]$, en de DEBAV van 18^h voor de randvoorwaarden op het interval $[18^h, 11^h]$ volgende dag].
3. De uurlijkse forecasts van $[00^h, 05^h]$ komen uit VAISALA runs die dagelijks om 18^h worden opgestart, met waargenomen wegdek- en dieptetemperatuur van 18^h , DEBAV 18^h voor de randvoorwaarden van $[18^h, 23^h]$, DEBAV 00^h voor de randvoorwaarden van $[00^h, 11^h]$.

³Deze DEBAV is op het moment van het opstarten van de run nog niet beschikbaar. De verschillen met de laatst beschikbare DEBAV zijn echter gering.

4. De uurlijkse forecasts van [06^h, 11^h] komen uit VAISALA runs die dagelijks om 00^h worden opgestart, met de waargenomen wegdek- en dieptetemperatuur van 00^h, DEBAV 00^h voor de randvoorwaarden van [00^h, 05^h], DEBAV 06^h voor de randvoorwaarden van [06^h, 11^h].

Merk op dat we in ons forecast klassificatie schema (1) en (2) kunnen samennemen.

In het Kalman-assimilatie en correctie schema, is voor ieder filter de begintoestand vast gekozen,

$$\hat{X}_0 = 0, \quad P_0 = \begin{pmatrix} 0.7 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{pmatrix}, \quad (3.1)$$

de ruis in de observatie vergelijking is tijdsafhankelijk, $r \equiv 0.25^\circ C$, en we verwaarlozen de ruis in de evolutie vergelijking, i.e. $Q(t) \equiv 0$. Dit betekent dat we in feite een lineaire regressie methode toepassen, een beperkt Kalman-filter, waarvan we de regressie coëfficiënten iteratief berekenen. Om het model ruimte te geven zich aan te passen aan de omgevingsomstandigheden, en te leren hoe de gemaakte forecasts relateren aan de waarnemingen, zijn de diagonaalelementen in de covariantiematrix P_0 relatief groot gekozen. Hierdoor worden gedurende de eerste iteratie stappen forse correcties toegestaan. De eerste maand aan gegevens wordt gebruikt als leerset om het filter op te leiden en de regressie coëfficiënten aan te passen aan de omstandigheden. Gedurende de volgende maanden worden de verificatiegrootheden als functie van de tijd bestudeerd. In een volledig Kalman-filter dient de covariantie $Q(t)$ van de systeemruis te worden gespecificeerd. Dit vereist een nauwkeurige afregeling van het filter. De verificatie periode die we tot onze beschikking hebben is daarvoor veel te kort. In deze periode is geen duidelijke seizoensvariatie aanwezig. Dit in tegenstelling tot het artikel van Persson [1990] waar een voorbeeld wordt gegeven van een situatie waar de sneeuwbedekking smelt door het systematisch toenemen van de temperaturen, en waar vervolgens de regressie coëfficiënten zich langzaam aanpassen aan de nieuwe situatie. Voorlopig beperken we ons dus alleen tot lineaire regressie, en is onze strategie om het model ieder nieuw winterseizoen tijdig opnieuw op te starten.

De fout in de verwachtingen wordt gedefiniëerd als $E(t) = T_{fc}(t) - T_{obs}(t)$, voor de VAISALA forecasts $T_{fc}(t)$, en $\hat{E}(t; t_f) = \hat{T}_{fc}(t; t_f) - T_{obs}(t)$, voor de Kalman-correcties $\hat{T}_{fc}(t; t_f)$. De Kalman-correcties hangen natuurlijk wel af van de lengte, t_f , van het voorinterval. De belangrijkste resultaten van de verificatie procedure kunnen nu als volgt worden samengevat:

- In de figuren (A.1) en (A.2) worden de tijdreeksen voor de fouten $E(t)$ en de correcties $\hat{E}(t; t_f)$ gepresenteerd voor een voorspelinterval van $t_f = 3$ uur. Opvallend is dat het Kalman-filter vooral in het begin van januari een aantal malen volledig de mist in gaat. De opgelegde correcties werken in deze gevallen alleen maar averechts. In de tweede helft van januari, en zeker in februari is deze trend bijna volledig verdwenen. Dit kan betekenen dat de systematische fouten in deze periode een ander karakter hebben en het filter te star is om zich aan te passen, of dat het filter begin januari nog niet volleerd is.
- Wat verder opvalt in de tijdreeksen zijn de extreme positieve temperatuurpieken, die regelmatig optreden na het middaguur. Deze worden veroorzaakt door een combinatie van het zich snel opwarmende wegdek, en het inschakelprobleem van het VAISALA model. In sommige gevallen is het Kalman-correctie schema zelfs in staat deze fouten met meer dan $5^\circ C$ te corrigeren. Maar de bereikte performance van het filter is toch nog veel te laag. De gemiddelde fout voor het gehele temperatuur bereik is weliswaar verdwenen, maar de standaarddeviatie, die rond het middaguur toch nog waarden van $1^\circ C$ bereikt, is veel te hoog. Het valt ook op dat het VAISALA model de negatieve wegdektemperaturen voor de avond en de nacht behoorlijk overschat. Deze trend is heel duidelijk aanwezig in de tweede helft van februari. Het Kalman-filter is echter uitstekend in staat om deze fouten tot een acceptabele waarde, een standaarddeviatie van ongeveer $0.7^\circ C$, te reduceren.
- In de figuren (A.3) tot en met (A.5) worden voor het voorspelinterval met lengte $t_f = 3$ uur, de scatterplots voor de kritische temperatuurzone $[-5^\circ C, +3^\circ C]$ gegeven. De presentatie is gesplitst in het dagdeel $[09^h, 17^h]$, de avond $[18^h, 23^h]$, en de nacht $[00^h, 08^h]$. Opvallend is hier de bijzonder slechte performance van het VAISALA model voor het dagdeel. De grote bias en standaarddeviatie wordt met name veroorzaakt door het inschakeleffect van het model. De toepassing van het Kalman-filter maakt van de VAISALA scatterplot in figuur (A.3), duidelijk een smalle en rechte band die naar de hoofddiagonaal is verschoven. De correcties zijn dus relatief gezien nogal groot, maar de performance is nog niet goed genoeg omdat de band van de tweede figuur in (A.3) niet smal genoeg is. Het Kalman-filter corrigeert verder de systematische onderschatte wegdektemperaturen gedurende de avond en de nacht. Dit is vooral duidelijk zichtbaar voor de negatieve wegdektemperaturen in het interval $[-5^\circ C, 0^\circ C]$, die gedurende de nacht gemiddeld zo'n $2^\circ C$ te laag zijn. Ook de band van de figuren (A.4) en (A.5) wordt door toepassen van het filter iets smaller. Wel dient opgemerkt te worden dat het VAISALA model redelijk goed presteert voor de avond en de nacht. Er is nog maar weinig scatter aanwezig, en de bias kun je makkelijk reduceren m.b.v. het filter tot hooguit een paar tiende graad.
- In de figuren (A.6) en (A.7) ziet U voor voorspelintervallen met een lengte van $t_f = 1$ uur tot en met $t_f = 6$ uur, de gemiddelde afwijking van de fout en de stan-

		00	03	06	09	15	18	21
VAISALA								
	ME	-1.2	-1.0	-0.8	-0.7	2.1	-1.3	-1.4
	SD	1.0	1.2	1.1	0.7	3.0	0.6	0.6
KALMAN								
1	ME	0.1	-0.2	0.0	0.1	0.1	0.2	0.0
	SD	0.6	0.5	0.5	0.5	1.0	0.5	0.6
2	ME	0.2	-0.1	-0.2	0.0	0.3	0.3	0.0
	SD	0.7	0.5	0.6	0.7	1.0	0.6	0.5
3	ME	0.2	-0.1	-0.1	0.1	0.8	0.3	0.2
	SD	0.7	0.6	0.6	0.7	1.0	0.5	0.6
4	ME	0.2	-0.1	-0.2	0.1	0.8	0.3	0.2
	SD	0.7	0.7	0.7	0.6	0.9	0.5	0.5
5	ME	0.1	-0.1	-0.1	0.0	1.0	0.2	0.2
	SD	0.7	0.7	0.7	0.6	0.9	0.6	0.5
6	ME	0.2	-0.1	-0.2	0.0	1.0	0.3	0.2
	SD	0.7	0.8	0.7	0.6	0.9	0.5	0.5

Tabel 3.1: Bias en standaarddeviatie als functie van t_f . Alleen voor het kritische temperatuur bereik $[-5^{\circ}C, +3^{\circ}C]$.

daarddeviatie van deze afwijking, voor de VAISALA- en de gecorrigeerde forecasts voor het gehele temperatuur bereik. Dezelfde resultaten, maar nu voor het kritische bereik van $[-5^{\circ}C, +3^{\circ}C]$, treft U aan in de figuren (A.8) en (A.9). De bias voor de gehele temperatuur range is na correctie verwaarloosbaar klein, zelfs voor verwachtingen tot 6 uur vooruit. Ook het dagdeel vertoont over het gehele temperatuur bereik weinig bias; echter, in het kritische temperatuur bereik is de bias na correctie duidelijk niet verwaarloosbaar. Met name gedurende de dag en vroege avond loopt deze op naar meer dan $0.5^{\circ}C$, met een maximum van $1^{\circ}C$ omstreeks 15^h 's middags, dit in geval $t_f = 6$ uur. De standaarddeviatie gedurende de avond en de nacht is in de orde van $0.5^{\circ}C - 0.7^{\circ}C$ over het gehele temperatuur bereik, en neemt slechts langzaam toe met de lengte t_f van het voorspelinterval. De winst die we hier boeken, t.o.v. de VAISALA forecasts, ligt in de orde van $0.5^{\circ}C$'s nachts en $0.2^{\circ}C$'s avonds. De performance overdag is uiteraard slecht, terwijl we hier toch nog de standaarddeviatie met ruim $1^{\circ}C$ kunnen reduceren. In het kritische temperatuur bereik hebben de correcties 's avonds en 's nachts ook weer een reductie van de standaarddeviatie tot gevolg van resp. $0.5^{\circ}C$ en $0.2^{\circ}C$. De correcties overdag leiden tot een reductie van de standaarddeviatie van bijna $2^{\circ}C$.

- In de figuren (A.10) en (A.11) worden voor de kritische temperaturen de bias en standaarddeviatie voor de tijdstippen 00^h , 03^h , 06^h , 09^h , 15^h , 18^h , en 21^h , nog eens extra gegeven als functie van de lengte t_f van het voorspelinterval. Deze resultaten vindt U ook numeriek terug in de tabel (3.1). Hieruit lezen we af dat de standaarddeviatie nauwelijks te reduceren is om 09^h 's ochtends, en 's avonds om 18^h en 21^h . Alleen de bias wordt hier fors gecorrigeerd.

Door een nauwkeurige analyse van het Kalman-filter, is het waarschijnlijk mogelijk deze resultaten op een aantal punten nog aan te scherpen. Dit kan bijvoorbeeld door de statistiek van het filter, met name $Q(t)$, geschikt te kiezen. Het is dan misschien zelfs mogelijk om het leerproces van het filter te beperken tot hooguit een tweetal weken.

4 De operationele opzet van het Kalman-filter

Het vervaardigen van de wegdektemperatuurvoorspellingen vormt een belangrijk onderdeel van de Weer en Verkeer (WenV) systemen van het KNMI. De WenV applicatie is gesplitst over een tweetal systemen, te weten **WenV1** en **WenV2**. Het **WenV1** systeem wordt ook wel **ICECAST** genoemd, en is operationeel van 1 november tot 1 april. In deze periode maakt de meteoroloog dagelijks interactief 4 wegdektemperatuurvoorspellingen. Deze worden via videotekst pagina's aan de klant ter beschikking gesteld. Het **WenV2** systeem wordt ook wel **AUTOCAST** genoemd.

Voor de beschrijving van de implementatie van het Kalman-filter, beperken we ons nu tot het **AUTOCAST** systeem. De KNMI medewerkers leveren tijdens het gebruik van **AUTOCAST** de DEBAV invoer. Daarnaast spelen ze natuurlijk een rol achter de schermen, dit i.v.m. onderhoud en het verwerken van klachten. Het systeem is operationeel van 1 november tot 1 april. Tijdens deze periode kunnen de klanten, en dat zijn de diverse dienstkringen van Rijkswaterstaat en de distrikten van de Provinciale Waterstaten, wegdekvoorspellingen laten maken. De klanten leveren hiervoor zelf de wegdekmeetgegevens in de vorm van gemeten oppervlakte- en dieptetemperaturen over een periode van 12 uur. In het systeem worden telkens de waarnemingen van de laatste 24 uur bijgehouden en gebruikt. Deze meetgegevens worden tijdens de berekening gecombineerd met de meteogegevens (de verwachtingen uit DEBAV) van het dichtsbijzijnde KNMI-meetstation. Als vervolgens alle voorbereidingen zijn getroffen, kan het VAISALA model worden gedraaid. Voordat de gemaakte wegdektemperatuurverwachting terug gaat naar de klant, wordt deze eerst gecorrigeerd met behulp van een Kalman-correctieschema.

In de huidige operationele opzet van het wegdektemperatuurverwachtingmodel, kan de klant op ieder gewenst tijdstip een wegdektemperatuurverwachting opvragen. Deze situatie houdt in dat het model op ieder uur van de dag kan worden opgestart, en elk

van deze modelruns heeft zijn eigen karakteristiek. Voor elke modelrun hebben we een apart Kalman-filter nodig dat dan vervolgens weer afhangt van het tijdstip waarvoor de verwachting wordt gemaakt, en van de lengte van het voorspelinterval. De eerste mogelijkheid sluiten we nu uit doordat we de modelruns fixeren tot een beperkt aantal; het model wordt in principe opgestart op vaste tijdstippen van de dag. De gebruiker merkt hier helemaal niets van. De gebruiker kan nog steeds op ieder gewenst tijdstip van de dag een verwachting opvragen. Als de gebruiker belt worden de laatste waarnemingen doorgestuurd. Vervolgens wordt het Kalman-correctie schema aangeroepen waar de laatst beschikbare VAISALA forecasts worden gecorrigeerd.

In de nieuwe operationele opzet wordt het VAISALA model 4 maal per dag opgestart, om precies te zijn om 12^h, 18^h, 00^h en 06^h, i.e. periodiek om de 6 uur. Het forecast klassificatie schema lijkt op het schema zoals beschreven in paragraaf 3. Het verschil is, dat de forecasts nu niet uit de vorige run worden gehaald. Uit een nader onderzoek is immers gebleken dat het inschakelverschijnsel alleen gedurende het dagdeel [09^h, 17^h] een rol speelt, en niet daarbuiten. Nadat de forecasts zijn gemaakt, worden de waargenomen wegdektemperaturen van de laatste 24 uur geassimileerd in het lineaire regressie model, en worden de regressie coëfficiënten automatisch bijgesteld. Voor deze periodieke actie is een afspraak gemaakt met de klant (RWS), zodat hij of zij deze actie zelf automatisch triggert. Het triggeren vindt plaats binnen 1 uur na de gevraagde tijden, i.e. voordat de waarneming van ($hh + 1$) uur binnen is, met $hh = 12, 18, 00, 06$. Tevens is er voor gezorgd dat de klanten om de beurt triggeren. Na het triggeren van het systeem, dient voor iedere gewenste lokatie een tijdcyclusschema te worden uitgevoerd, waarbij de specifieke data voor de lokatie moet worden opgevraagd. Alvorens we dit schema presenteren, eerst een aantal definities van datum, uur, en de files die we gebruiken om de forecasts, waarnemingen, en de filterresultaten weg te schrijven:

Definities

dtm de datum uitgedrukt in jaar, maand en dag, i.e. **dtm** = yymmdd.

nextdtm de datum van de dag die volgt op **dtm**.

prevdtm de datum van de dag die vooraf gaat aan **dtm**.

hh het uur van de dag.

#{dtm}hh.fct de file die de VAISALA forecasts bevat van de run die is opgestart om **hh** uur, met **hh** = 00, 06, 12, en 18.

#{dtm}hh.obs de file die de laatste 24 uur aan waarnemingen voor de wegdektemperatuur bevat, met **hh** = 00, 06, 12, en 18.

$\${dtm}hh.res$ de restart file van het filter voor de dag dtm en het uur hh , met $hh = 00, 06, 12, \text{ en } 18$.

$\${dtm}hh.reg$ de file die de regressie coëfficiënten bevat voor de forecasts uit de file $\${dtm}hh.fct$, nadat de laatste waarnemingen uit de file $\${nextdtm}hh.obs$ zijn geassimileerd.

$\${dtm}th.his$ de file die de geschiedenis van de waargenomen wegdektemperaturen bevat van de laatste 24 uur, met als laatste de waarneming op het actuele uur th .

$\${dtm}th.cor$ de file die de gecorrigeerde forecasts bevat waarbij op het uur th de laatste waarneming is binnengekomen.

Voor het format van de gebruikte files verwijzen we naar appendix B.

Iedere klant dient zich nu aan een viertal basisspelregels te houden, wanneer hij of zij wegdektemperatuurvoorspellingen laat maken:

1. Iedere klant start het systeem dagelijks **automatisch** op om $00^h, 06^h, 12^h, \text{ en } 18^h$.
2. Alle klanten **triggeren** het systeem **om de beurt**. Het systeem kan immers maar één klant tegelijkertijd verwerken. De reden hiervan is dat er slechts één modem is. Bij voorkeur krijgt iedere klant een triggerinterval van 2 of 3 minuten.
3. Als de klant het systeem automatisch opstart, dan hoeft hij niet meer te wachten op het draaien van het VAISALA model. Het maken van de modelforecasts gebeurt achter de schermen. De klant krijgt echter wel een gecorrigeerde forecast terug, op basis van de laatst gemaakte modelrun precies 6 uur geleden, en de laatste waarneming op het actuele tijdstip waarop de klant het AUTOCASST systeem heeft opgestart. Het voordeel van deze aanpak is dat het triggerinterval kan worden beperkt tot hooguit een paar minuten.
4. Natuurlijk kan de klant op ieder gewenst tijdstip van de dag handmatig een verwachting opvragen. Het VAISALA model wordt dan echter niet opgestart, maar de meest recente VAISALA forecasts worden gecorrigeerd m.b.v. het Kalman-correctie schema, waarbij tevens de laatste waarnemingen worden gebruikt, en de gecorrigeerde verwachtingen gaan vervolgens terug naar de klant.

Na het automatisch triggeren van het systeem, worden de volgende acties en controles uitgevoerd, afhankelijk van het tijdstip $hh \in \{00, 06, 12, 18\}$ van de dag.

Het automatische tijdcyclusschema

(A) De automatische productie van forecasts

- Controleer of de forecastfile $\{\text{dtm}\}\text{hh.fct}$ aanwezig is.
 - (Ja) Ga naar de Kalman-correctie module.
 - (Nee) (1) Geef de klant een gecorrigeerde modelforecast die is gebaseerd op de modelrun van 6 uur geleden, en de laatste waarneming die de klant zelf heeft gestuurd. Ga hiervoor naar de Kalman-correctie module. De klant logt nu uit en de telefoonlijn is weer vrij voor de volgende klant. Het maken van de nieuwe modelforecasts van hh uur, vindt bij het KNMI plaats, i.e. achter de schermen.
 - (2) Kijk of de juiste DEBAV er is en de wegdekmeetgegevens. Controleer de invoer van het VAISALA model. Indien deze correct is, start dan het VAISALA programma op met de wegdekmetingen van hh uur. Er wordt nu een 24-uurs verwachting gemaakt. Indien nodig wordt hiervoor de dagovergang om 12 uur overlapt, door de run vanaf dat tijdstip weer opnieuw op te starten. Hou de status van het VAISALA programma bij. Als de run goed gaat, dan wordt de uitvoer in de file `modelout.dat` geschreven. Dit is een ringbuffer waarin de gemaakte forecasts in een bepaald format zijn opgeslagen. Deze forecasts worden hieruit gehaald en weggeschreven naar de file $\{\text{dtm}\}\text{hh.fct}$.
Ga nu naar de Kalman-assimilatie module.

(B) De Kalman-assimilatie module

In deze module assimileren we de waarnemingen van de laatste 24 uur in het lineaire regressie model. Hiervoor worden de forecasts uit de file $\{\text{prevd}\text{tm}\}\text{hh.fct}$ gebruikt. Het programma dat deze actie uitvoert heet **kalassim**, het is geschreven in fortran.

- Controleer of de file met forecasts $\{\text{prevd}\text{tm}\}\text{hh.fct}$, en de file met waarnemingen $\{\text{dtm}\}\text{hh.obs}$ in het juiste format aanwezig zijn. De restart file voor het filter $\{\text{prevd}\text{tm}\}\text{hh.res}$ is altijd aanwezig.
 - (Ja) Assimileer de data door het programma **kalassim** op te starten met als invoer de files
 - (*) de VAISALA forecast file $\{\text{prevd}\text{tm}\}\text{hh.fct}$,
 - (*) de waargenomen wegdektemperaturen in de file $\{\text{dtm}\}\text{hh.obs}$,
 - (*) de filter restart file $\{\text{prevd}\text{tm}\}\text{hh.res}$.

Op elk moment houden we de status van het programma **kalassim** bij. Als de run succesvol is, dan schrijven we de regressie coëfficiënten naar de file $\{\text{prevdtm}\}\text{hh.reg}$, en het restart filter voor de volgende dag naar de file $\{\text{dtm}\}\text{hh.res}$. Als de run niet succesvol is, dan hergebruiken we de laatste restart file $\{\text{prevdtm}\}\text{hh.res}$, en de file met regressie coëfficiënten $\{\text{prevprevdtm}\}\text{hh.reg}$.

Ga nu naar de Kalman-correctie module.

(Nee) Data-assimilatie is nu niet mogelijk, maar neem nu voor de restart file, en de file met de regressie coëfficiënten, de voorgangers $\{\text{prevdtm}\}\text{hh.res}$ en $\{\text{prevprevdtm}\}\text{hh.reg}$, i.e.

$$\{\text{prevdtm}\}\text{hh.res} \rightarrow \{\text{dtm}\}\text{hh.res},$$

$$\{\text{prevprevdtm}\}\text{hh.reg} \rightarrow \{\text{prevdtm}\}\text{hh.reg}.$$

Ga nu naar de Kalman-correctie module.

De Kalman-correctie module

Het corrigeren van de VAISALA forecasts wordt uitgevoerd door het fortran programma **kalcorr**. Voordat deze routine wordt aangeroepen bepalen we eerst het actuele tijdstip t_a in gehele uren, en de bijbehorende datum dtm . Vervolgens controleren we of de wegdekmeting op het tijdstip t_a , i.e. $tobs(t_a)$, aanwezig is.

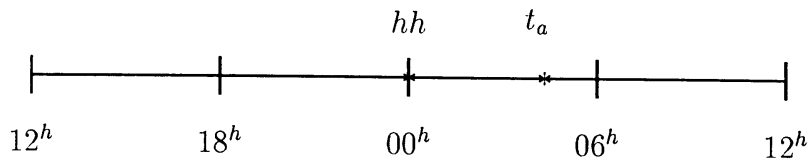
(Ja) We bepalen nu de bijbehorende forecastfile $\{\text{dtm}\}\text{hh.fct}$, zie ook figuur (4.1).

- Als $t_a \in [12^h, 18^h)$, dan $hh = 12$,
- Als $t_a \in [18^h, 00^h)$, dan $hh = 18$,
- Als $t_a \in [00^h, 06^h)$, dan $hh = 00$,
- Als $t_a \in [06^h, 12^h)$, dan $hh = 06$.

Controleer of de file $\{\text{dtm}\}\text{hh.fct}$ aanwezig is.

(Ja) Controleer of de file met regressie coëfficiënten $\{\text{prevdtm}\}\text{hh.reg}$ aanwezig is. Zo ja, dan wordt het Kalman-correctie programma **kalcorr** aangeroepen met als invoer de forecasts in $\{\text{dtm}\}\text{hh.fct}$, de laatste waarneming $tobs(t_a)$, en de regressie coëfficiënten in de file $\{\text{prevdtm}\}\text{hh.reg}$. Hou de status van het programma **kalcorr** bij. Als de run succesvol is, dan worden de gecorrigeerde forecasts in de file $\{\text{dtm}\}\text{ta.cor}$ gezet, en deze gaat naar de klant. Bij een crash krijgt de klant de ongecorrigeerde VAISALA verwachting uit de file $\{\text{dtm}\}\text{hh.fct}$.

- (Nee) De VAISALA forecastfile $\{dtm\}hh.fct$ is niet aanwezig. Er wordt nu analoog als hierboven een verwachting gemaakt op basis van de forecastfile $\{dtm\}(hh-6).fct$, indien deze aanwezig is. Als ook deze niet aanwezig is, kan er geen verwachting worden gemaakt, en krijgt de klant een foutmelding.
- (Nee) De wegdekmeting van t_a uur is niet aanwezig. Er kan nu geen verwachting worden gemaakt, en de klant krijgt een foutmelding.



Figuur 4.1: Het actuele tijdstip t_a en het tijdstip hh waarop de laatste VAISALA forecast is gemaakt.

Buiten de modules die hierboven zijn beschreven, en welke automatisch worden aangeroepen op vaste tijden van de dag, kan de klant verder op elk gewenst tijdstip een gecorrigeerde verwachting opvragen. In dit geval wordt alleen de Kalman-correctie module zoals boven beschreven aangeroepen.

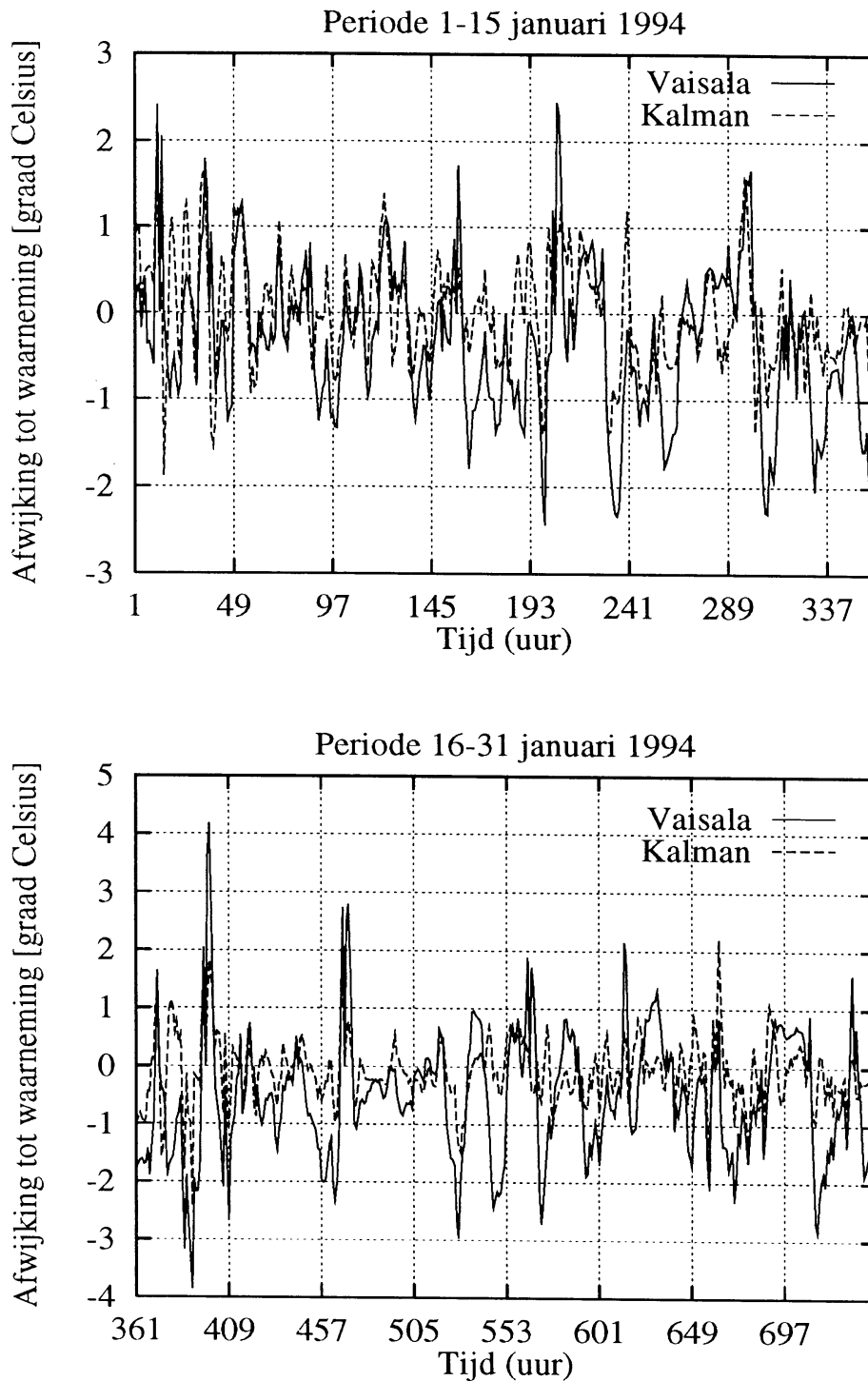
Dankwoord

Op deze plaats wil ik Jeanette Onvlee, Seijo Kruizinga, Kees Kok, Dick Riepma, en Henny Vergouw bedanken voor het aandachtig lezen, advies en opmerkingen op de eerste versie van deze rapportage. Daarnaast wil ik Herman Wessels bedanken voor zijn advies aangaande het onderzoek dat heeft geleid tot deze rapportage. Tenslotte dank ik Nader Emami Salut voor het maken van de VAISALA forecasts en voor de uitleg van de werking van het oude operationele wegdektemperatuurvoorspelsysteem.

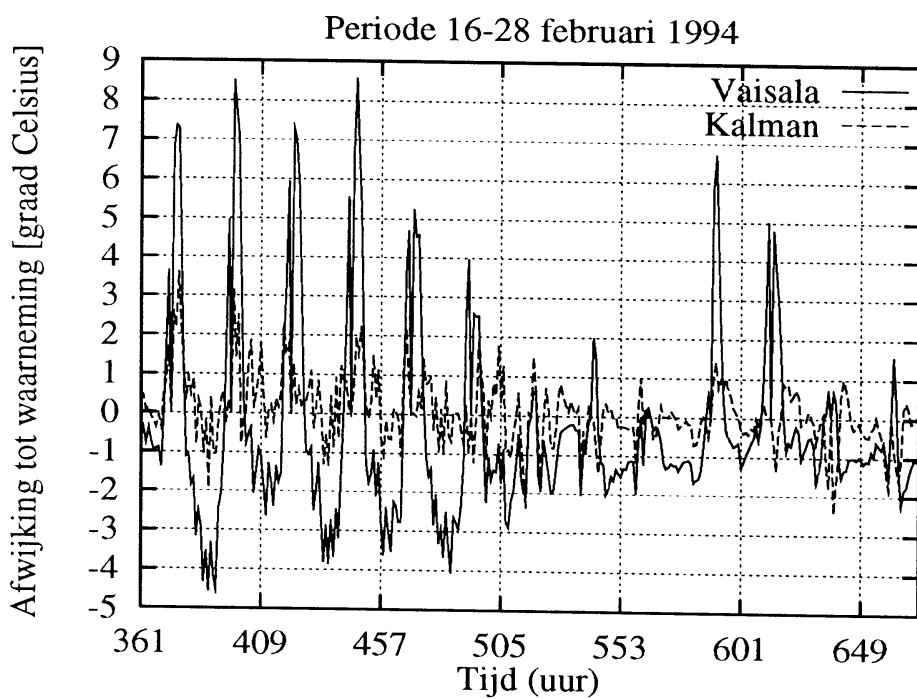
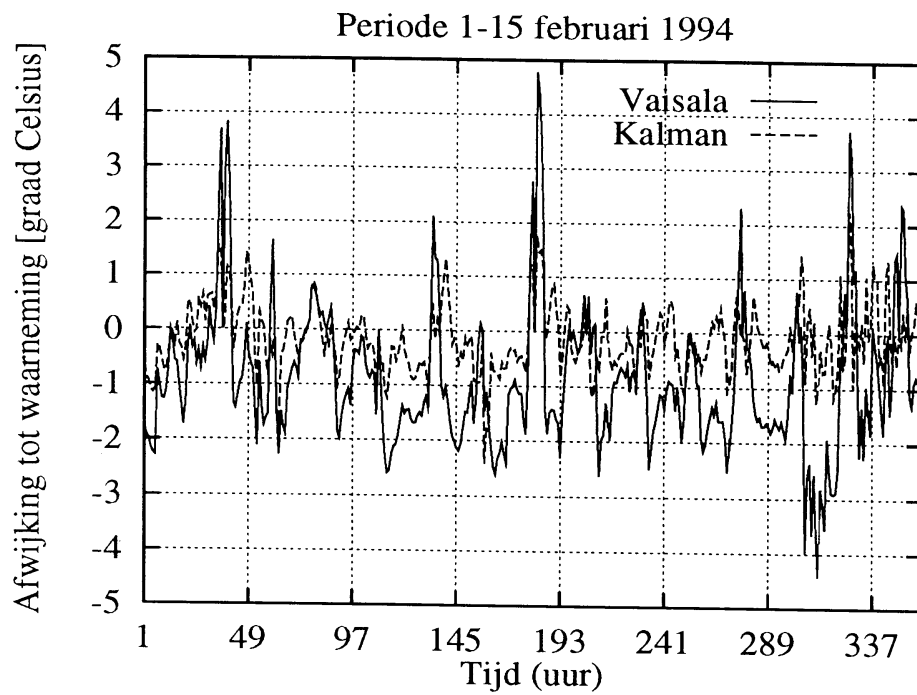
Referenties

- [1] Bryson, A.E., and Y.-C. Ho, *Applied optimal control; optimization, estimation, and control*.
Hemisphere, New York, 1975.
- [2] Jansen, S.T.H., *Beoordeling van twee modellen voor het voorspellen van de temperatuur van het wegdek in het kader van het Weer & Verkeer project*.
TNO-rapport, projectnummer: 741410392, TNO Wegtransportmiddelen Delft, 1994.
- [3] Kok, C.J., en C.L.M. van Vliet, *Gids ten behoeve van de weersverwachting tijdens de WK94*
Memorandum WM-94-13 (ongepubliceerd manuscript), KNMI, De Bilt, 1994.
- [4] Maybeck, P.S., *Stochastic models, estimation, and control, Vol. I*.
Academic Press, New York, 1979.
- [5] Minnaert, M., *De natuurkunde van 't vrije veld: Deel 2: Geluid, warmte, elektriciteit*.
Thieme - Zutphen, 1975.
- [6] Persson, A., *Kalman filtering - a new approach to adaptive statistical interpretation of numerical meteorological forecasts*.
ECMWF newsletter, no. 46 (1989); p. 16-20.
- [7] Rayner, P.J., *The Meteorological Office Forecast Road Surface Temperature Model*.
The meteorological magazine, Vol. 116, no. 1379, 1987, p. 180-191.
- [8] Thornes, J.E., *The prediction of ice formation on motorways*.
Ph.D. thesis, University of London, 1984.

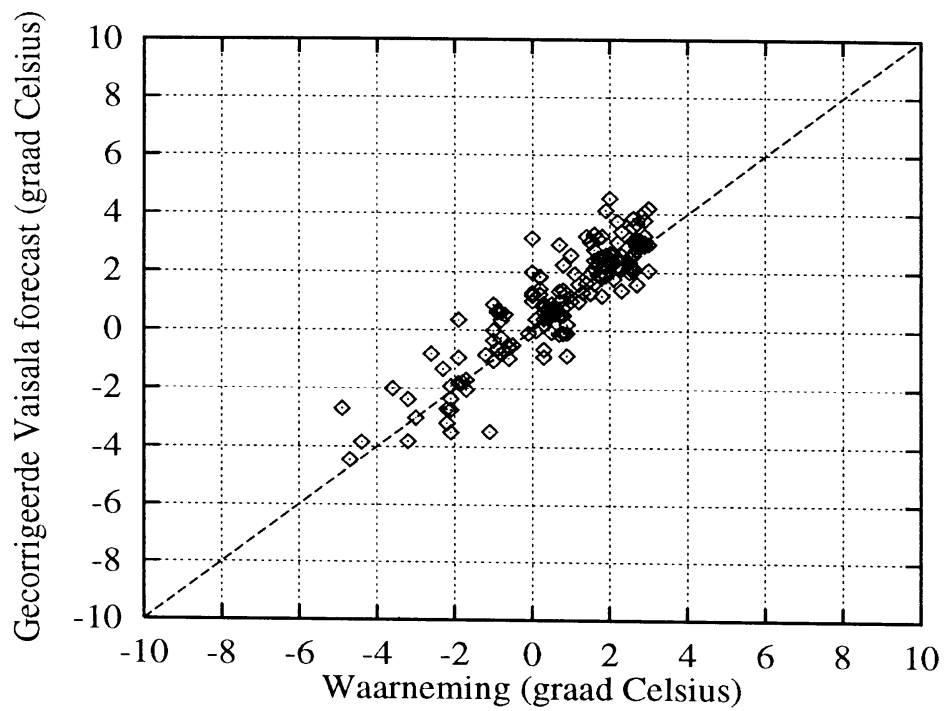
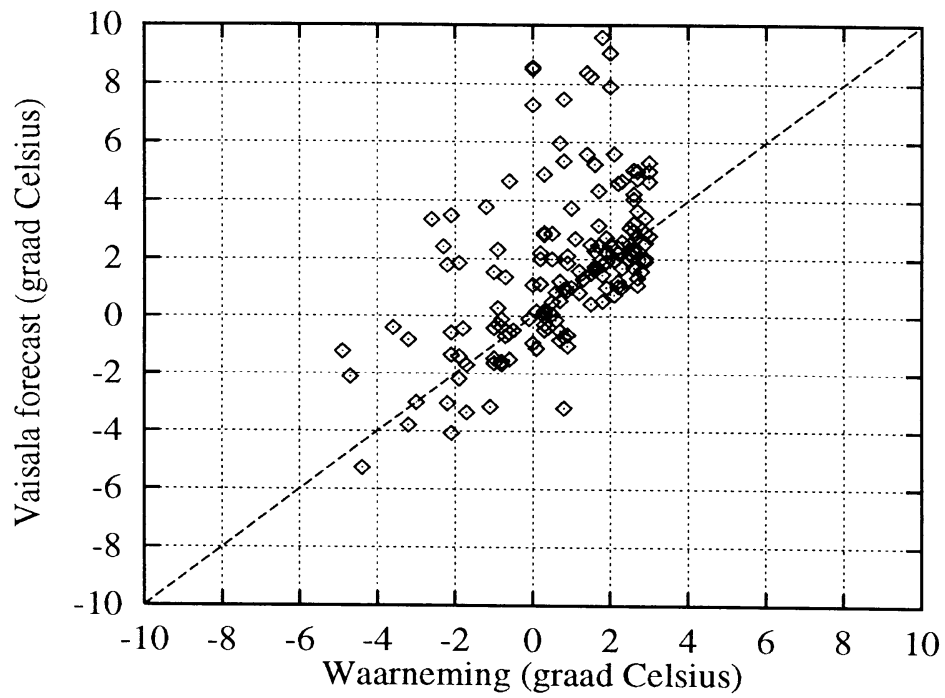
A Grafische visualisatie van de verificatie resultaten



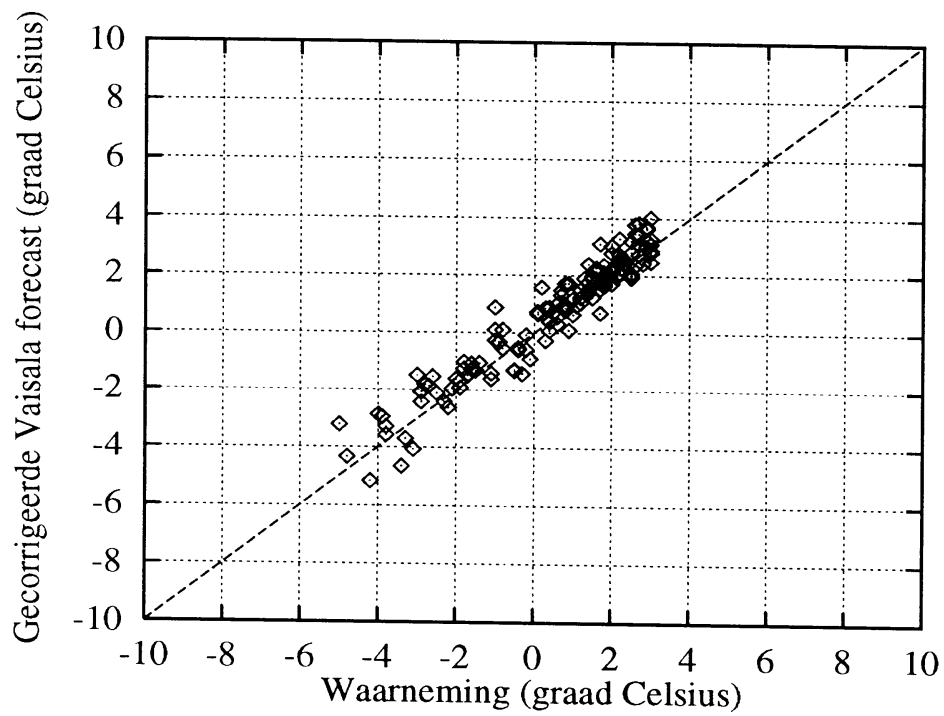
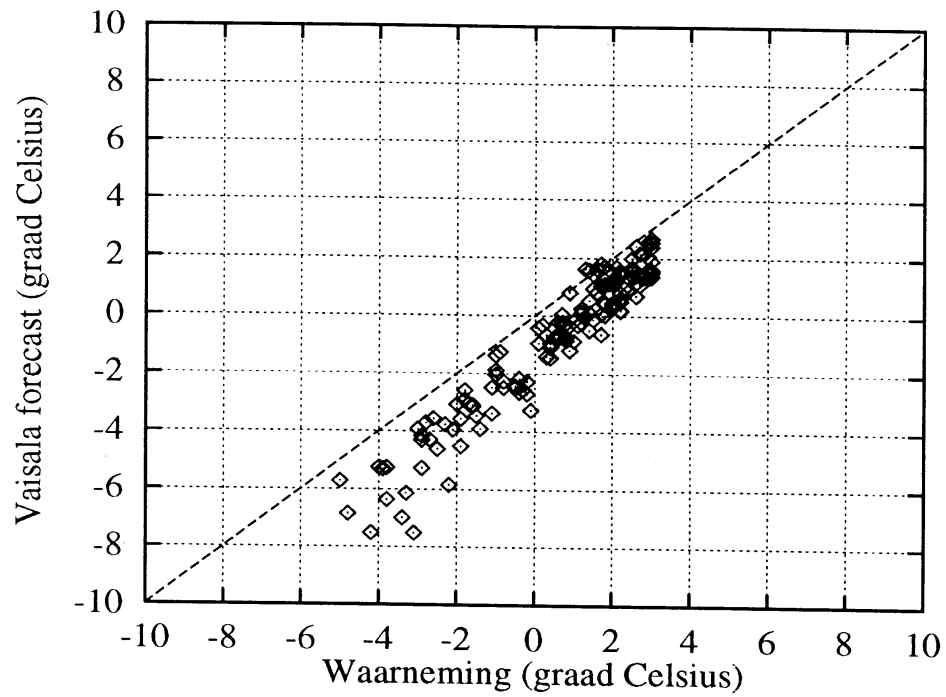
Figuur A.1: De fout in de VAISALA- en de gecorrigeerde verwachtingen.



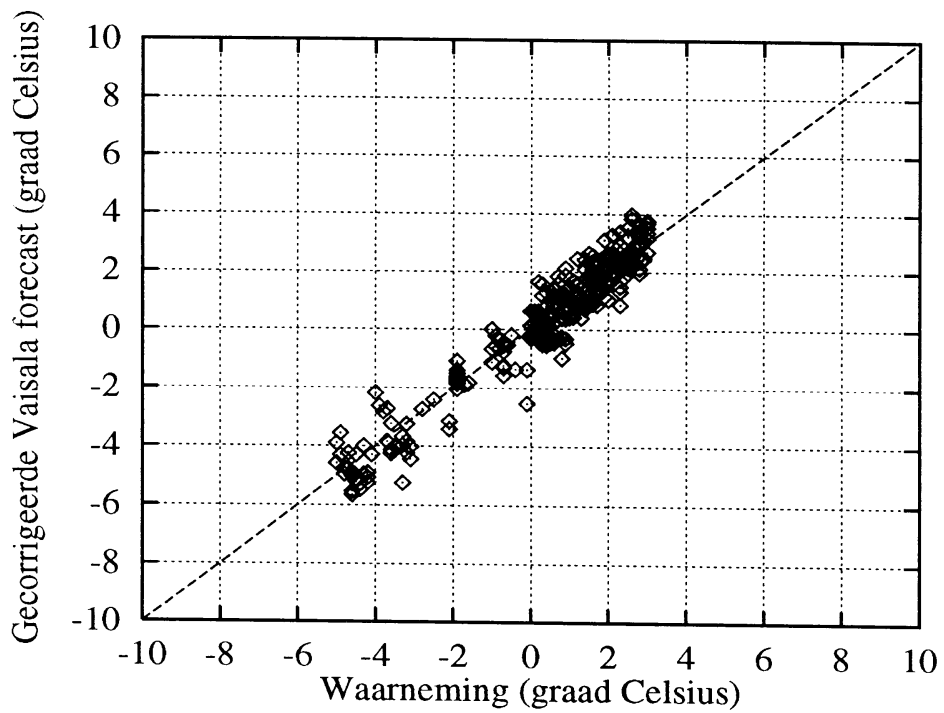
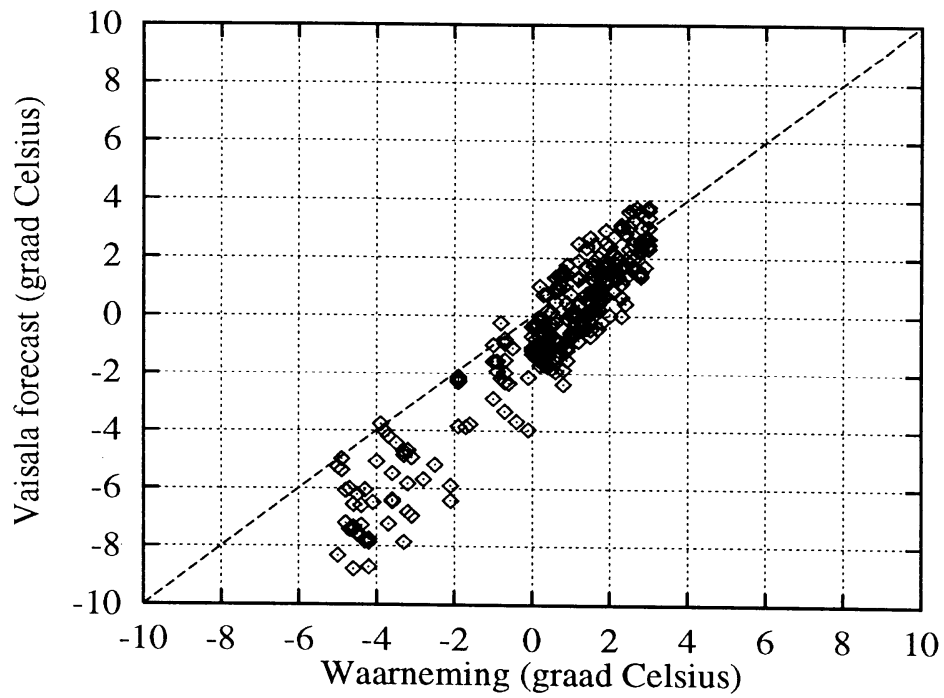
Figuur A.2: De fout in de VAISALA- en de gecorrigeerde verwachtingen.



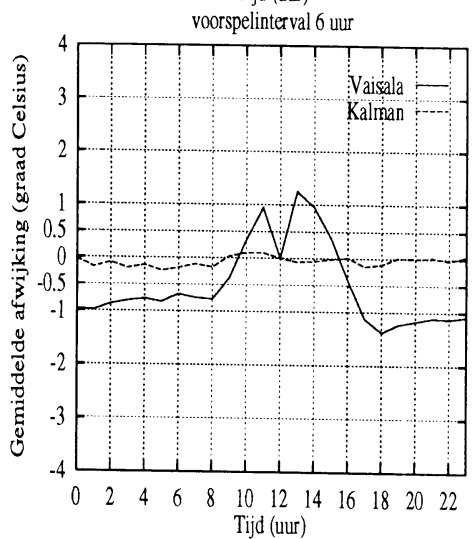
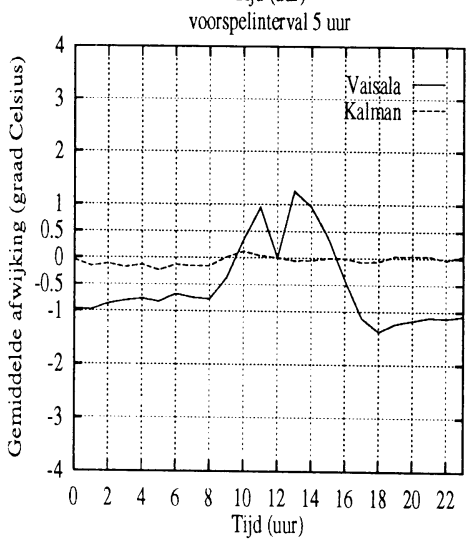
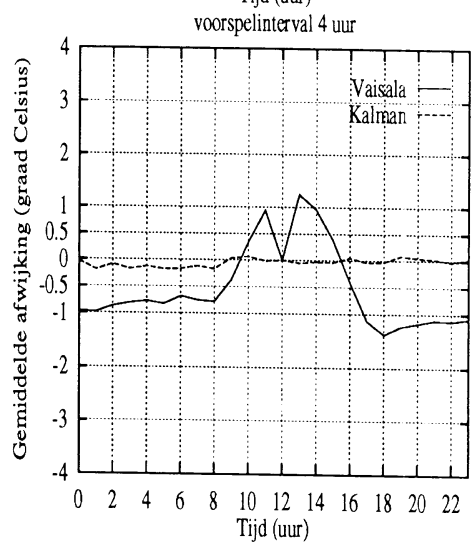
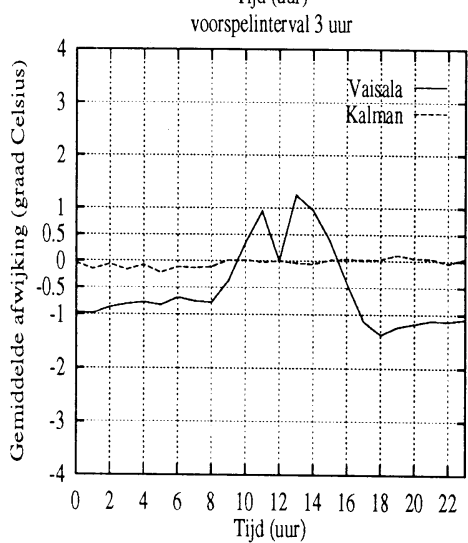
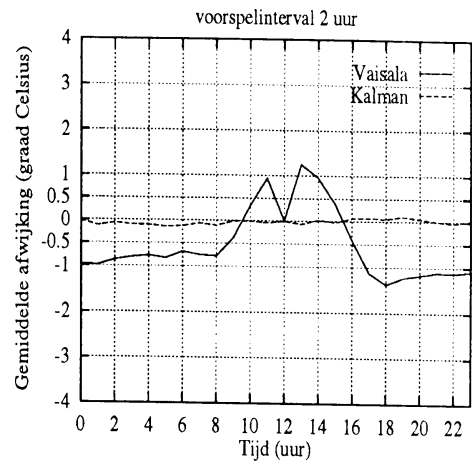
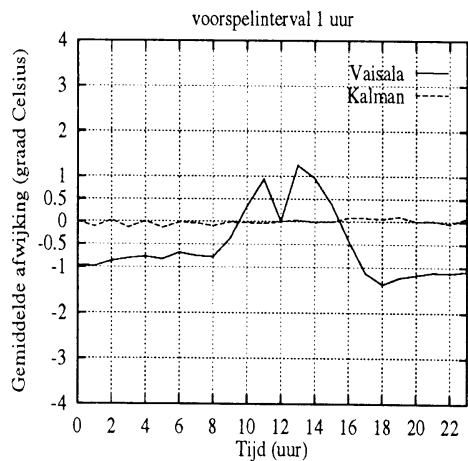
Figuur A.3: De scatter-plots voor het dagdeel $[09^h, 17^h]$.



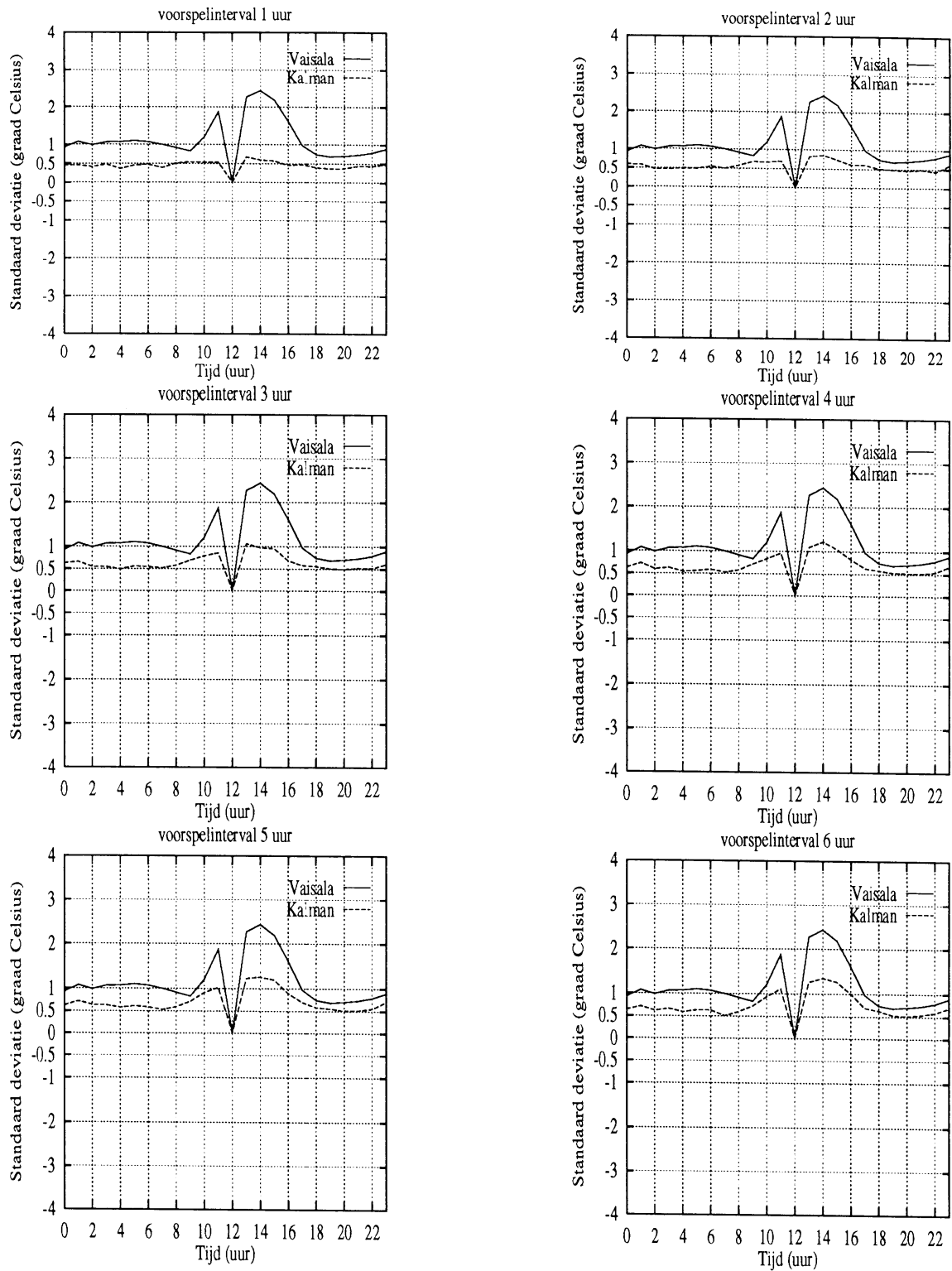
Figuur A.4: De scatter-plots voor de avond $[18^h, 23^h]$.



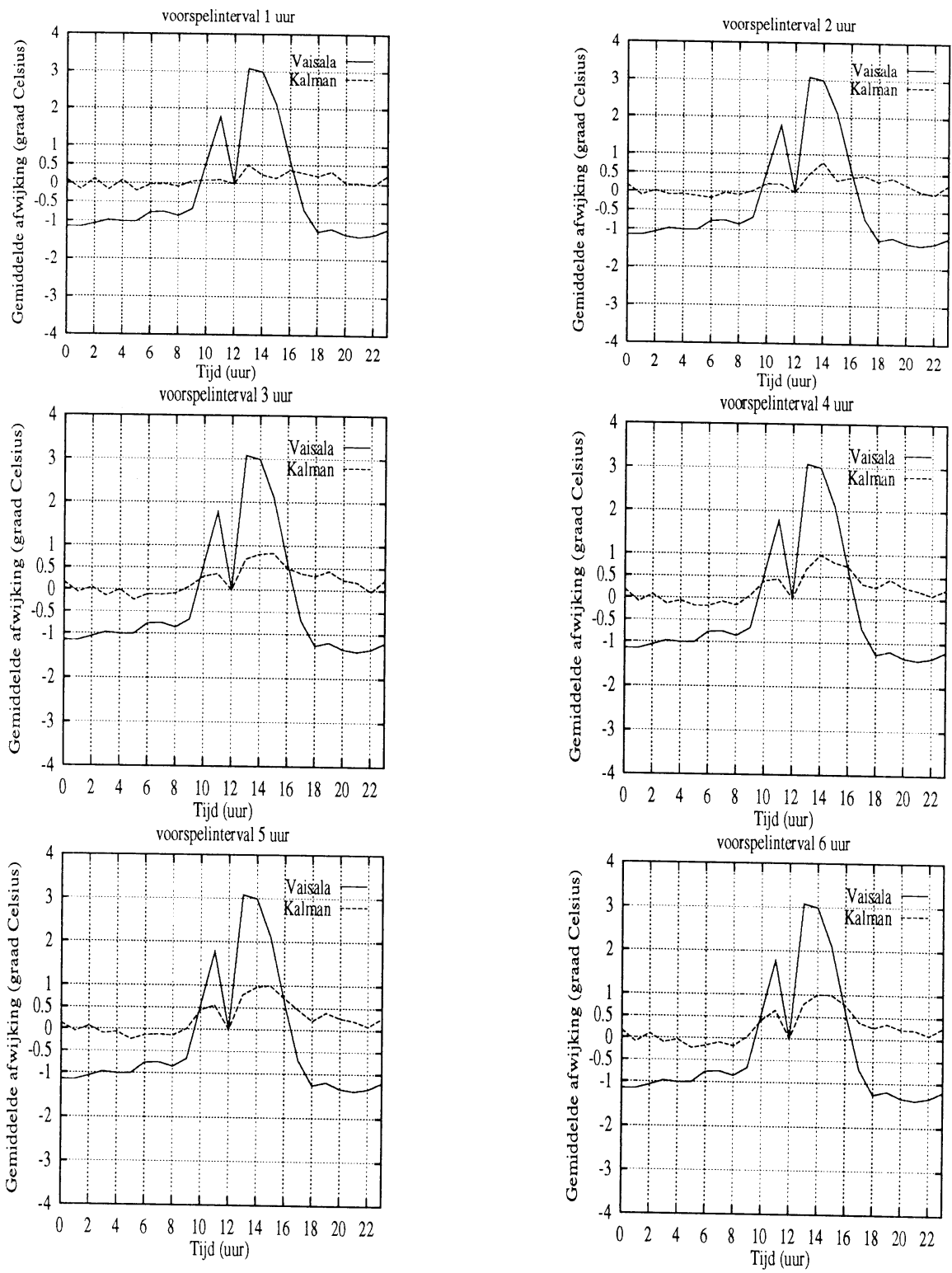
Figuur A.5: De scatter-plots voor de nacht en de vroege ochtend [00^h, 08^h].



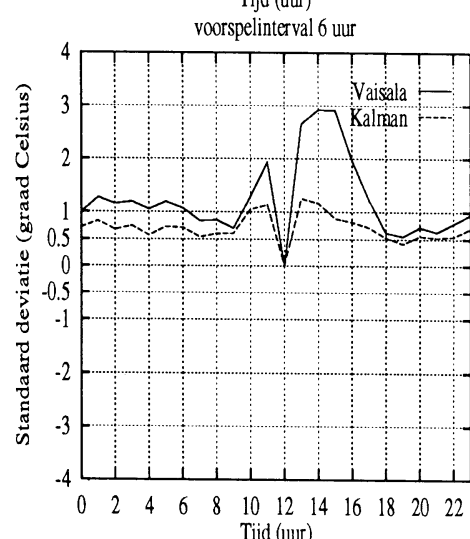
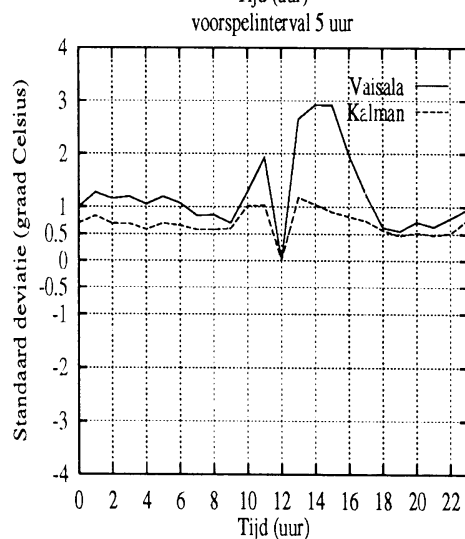
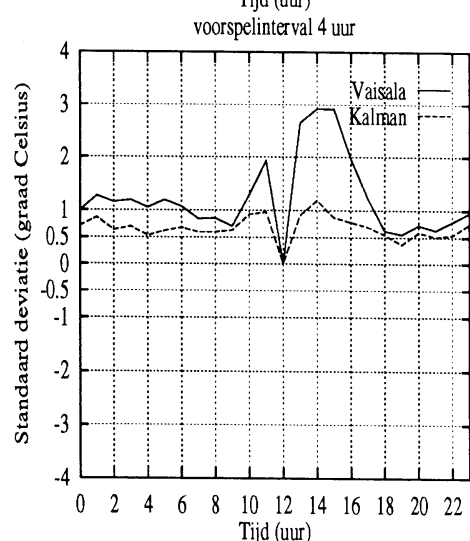
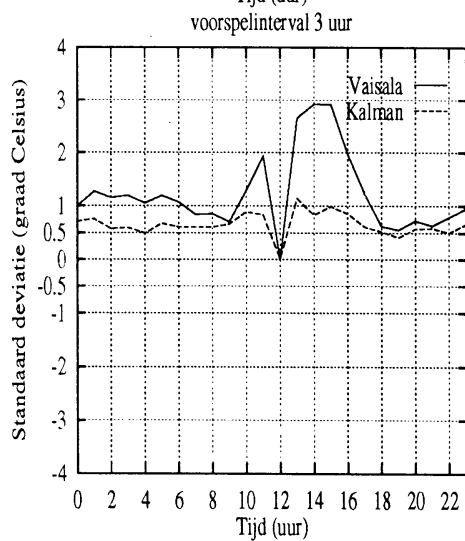
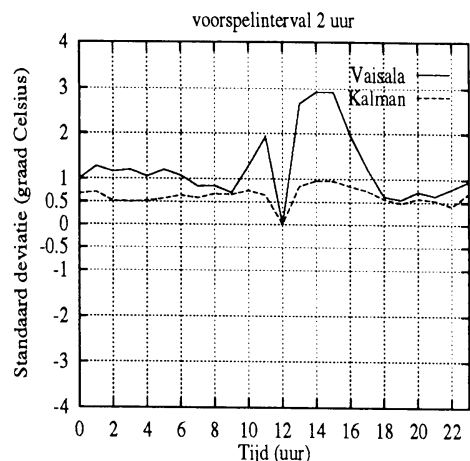
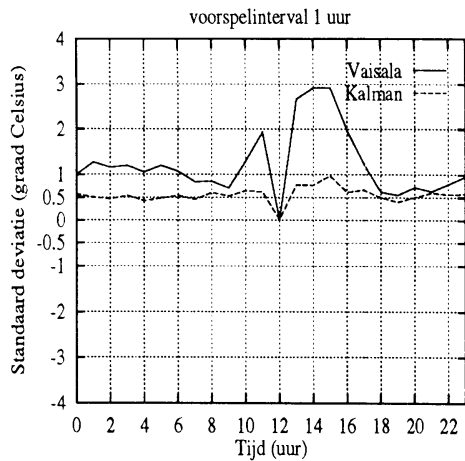
Figuur A.6: De gemiddelde afwijking in de VAISALA- en de gecorrigeerde verwachtingen (het gehele temperatuur bereik).



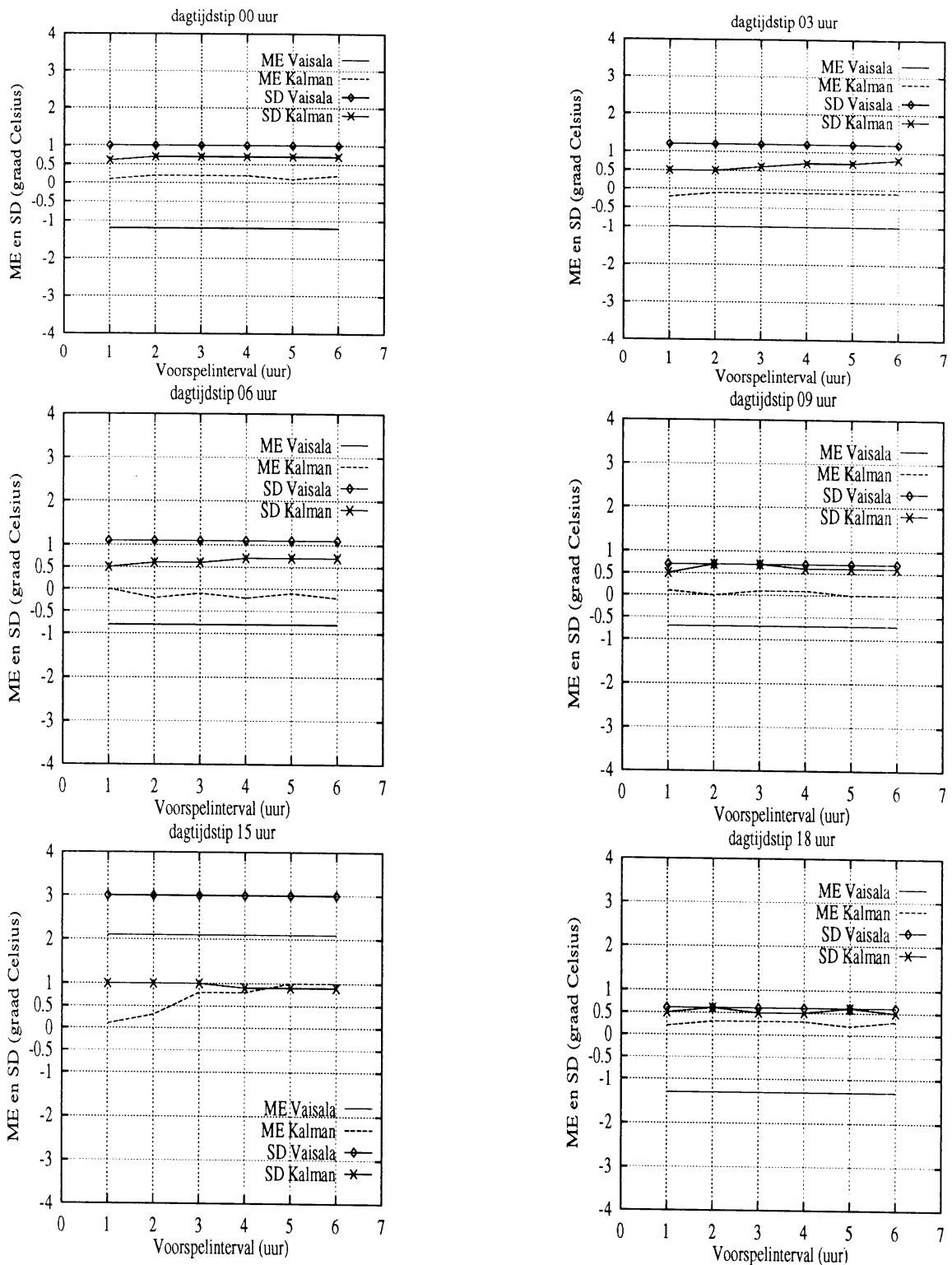
Figuur A.7: De standaard deviatie in de VAISALA- en de gecorrigeerde verwachtingen (het gehele temperatuur bereik).



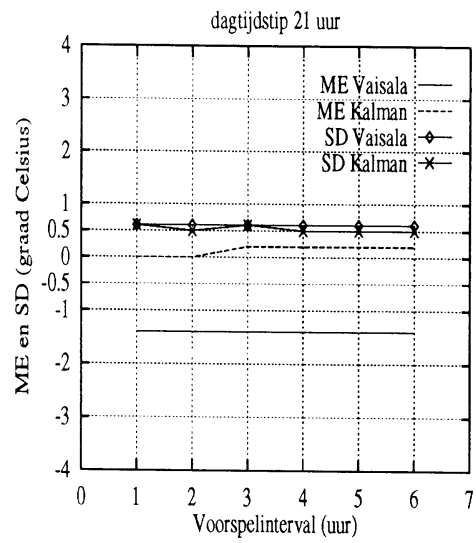
Figuur A.8: De gemiddelde afwijking in de VAISALA- en de gecorrigeerde verwachtingen (het kritieke temperatuur bereik $[-5^{\circ}C, +3^{\circ}C]$).



Figuur A.9: De standaard deviatie in de VAISALA- en de gecorrigeerde verwachtingen (het kritieke temperatuur bereik $[-5^{\circ}C, +3^{\circ}C]$).



Figuur A.10: De gemiddelde afwijking (ME) en standaard deviatie (SD) van de VAISALA- en de gecorrigeerde forecasts, als functie van het voorspelinterval (het kritieke temperatuur bereik $[-5^{\circ}C, +3^{\circ}C]$). 33



Figuur A.11: De gemiddelde afwijking (ME) en standaard deviatie (SD) van de VAISALA- en de gecorrigeerde forecasts, als functie van het voorspelinterval (het kritieke temperatuur bereik $[-5^{\circ}C, +3^{\circ}C]$).

B Het format van de input/output van de Kalman programmatuur

De Kalman programmatuur bestaat uit een tweetal fortran programma's, om precies te zijn **kalassim** en **kalcorr**.

Het programma **kalassim** leest het tijdstip **tv** waarop de gebruikte forecastfile is gemaakt, $t_v \in \{00, 06, 12, 18\}$. In het programma worden de waarnemingen uit de file **obs.dat** geassimileerd in het lineaire regressie model. Hiervoor worden de forecasts uit de file **fct.dat** gebruikt. Het Kalman-filter wordt voor deze aktie opgestart met initiële condities die opgeslagen zijn in de file **filter.ini**. Na assimilatie worden de nieuwe regressie coëfficiënten weggeschreven naar de file **filter.reg**, en het restart filter voor de volgende run naar de file **filter.res**.

Als de actuele datum **dtm** is, dan geldt:

input $\{\text{prevdtm}\}\text{tv.fct} \rightarrow \mathbf{fct.dat}$
 $\{\text{dtm}\}\text{tv.obs} \rightarrow \mathbf{obs.dat}$
 $\{\text{prevdtm}\}\text{tv.res} \rightarrow \mathbf{filter.ini}$

output $\mathbf{filter.reg} \rightarrow \{\text{prevdtm}\}\text{tv.reg}$
 $\mathbf{filter.res} \rightarrow \{\text{dtm}\}\text{tv.res}$

De file $\{\text{prevdtm}\}\text{tv.fct}$ bevat 7 kolommen met elk 24 regels. In de kolommen staat resp.

nrstat nrsens yy mm dd hh tfct,

in het fortran format **(I4,5(1X,I2),1X,F6.2)**. Hier is **nrstat** het nummer van het RWS station waar de waarnemingen vandaan komen, **nrsens** het sensornummer, **yy** het jaar, **mm** de maand, **dd** de dag, **hh** het uur, en **tfct** de VAISALA forecast. De datumtijdgroep, **dtg=yymmddhh**, in deze file begint met **dtg={prevdtm}tv**.

De file $\{\text{dtm}\}\text{tv.obs}$ bevat tevens 7 kolommen met elk 24 regels. In de kolommen staat resp.

nrstat nrsens yy mm dd hh tobs,

in het fortran format **(I4,5(1X,I2),1X,F6.2)**. Hier is **nrstat** het nummer van het RWS station waar de waarnemingen vandaan komen, **nrsens** het sensornummer, **yy** het jaar, **mm** de maand, **dd** de dag, **hh** het uur, en **tobs** de waargenomen wegdektemperatuur. De datumtijdgroep, **dtg=yymmddhh**, in deze file begint met **dtg={prevdtm}tv**.

Voor een nadere specificatie van het format van de files $\{\text{prevdtm}\}\text{tv.reg}$, en $\{\text{dtm}\}\text{tv.res}$, herinneren we de lezer aan de Kalman-correctie formule

$$\hat{T}_{\text{fc}}(t_a + t_f; t_f) = \hat{X}_0(t_a + t_f - 24^h; t_f) + \hat{X}_1(t_a + t_f - 24^h; t_f) \times T_{\text{fc}}(t_a + t_f) + \hat{X}_2(t_a + t_f - 24^h; t_f) \times T_{\text{obs}}(t_a), \quad (\text{B.1})$$

met $t_f = 0, 1, 2, \dots$, de lengte van het voorspelinterval, en t_a het aktuele tijdstip waarop de klant belt en om een verwachting vraagt. We hebben voor deze notatie gekozen omdat de regressie coëfficiënten behalve van t_a ook een functie van t_f zijn.

We variëren nu t_a van $t_a = t_v$ tot en met $t_a = t_v + 11$ (zie ook paragraaf (4) en figuur (4.1)), in stappen van 1 uur, en t_f van $t_f = 0$ tot en met $t_f = t_f^{\text{max}}$, in stappen van 1 uur, met $t_f^{\text{max}} = 23 - \text{index}(t_a)$. Hier is $\text{index}(t_a)$ de positie van het uur t_a t.o.v. t_v , i.e.

$$\text{index}(t_a) = \begin{cases} t_a - t_v, & \text{als } t_a - t_v \geq 0, \\ t_a + 24 - t_v, & \text{als } t_a - t_v < 0. \end{cases} \quad (\text{B.2})$$

Voor $t_v = 12^h$ vinden we dan:

$$\begin{array}{ll} t_a = 12^h, & t_f = 0^h, 1^h, \dots, 23^h, \\ t_a = 13^h, & t_f = 0^h, 1^h, \dots, 22^h, \\ \dots\dots, & \dots\dots\dots\dots\dots\dots, \\ t_a = 23^h, & t_f = 0^h, 1^h, \dots, 12^h. \end{array}$$

Omdat ieder filter 3 regressie coëfficiënten nodig heeft, levert ons deze aanpak in totaal op, $3 \times \sum_{k=13}^{24} k = 666$ regressie coëfficiënten voor 222 filters. Er zijn dus ook 222 restart filters na iedere run. Dit aantal wordt minder als we de maximale voorspeltermijn beperken tot 6 of 12 uur. Bij een maximale voorspeltermijn van 12 uur hebben we 432 regressie coëfficiënten voor 144 filters. Voor 6 uur bedragen deze waarden resp. 216 regressie coëfficiënten en 72 filters.

Bij de implementatie van het Kalman-filter hebben we gekozen voor de maximaal haalbare voorspeltermijn. We hebben de restart filters voor $t_a \in \{0, 1, \dots, 11\}$ en $t_f \in \{0, 1, \dots, 23\}$ in één file geplaatst. Iedere waarde van t_a correspondeert met een kolom uit die file, met variabele lengte $t_f^{\text{max}} + 1$. De files voor de regressie coëfficiënten en de restart filters zijn binair. De restart file $\{\text{dtm}\}\text{tv.res}$ voor het filter heeft het volgende format:

$$t_f \rightarrow \begin{cases} t_a = t_v & t_a = t_v + 1 & \dots\dots & t_a = t_v + 11 \\ \hat{X}_0, \hat{X}_1, \hat{X}_2 \\ P_0 \\ Q \\ r \end{cases} \quad (\text{B.3})$$

en dit blok herhaalt zich voor iedere $t_f = 0^h, 1^h, \dots, 23^h$. Als we hierin de covariantie matrices P_0, Q , en de covariantie r , weglaten, dan is dit precies het format van de file $\{\text{prevdtm}\}\text{tv.reg}$.

Het programma **kalcorr** leest het tijdstip **tv** waarop de gebruikte forecastfile is gemaakt, en het tijdstip **ta** waarop de klant belt, en dus de laatste waarneming binnen is. In het programma worden de forecasts uit de file **fct.dat** gecorrigeerd met behulp van de waarneming op het actuele uur t_a , i.e. $\text{tobs}(t_a)$, die in de file **obs.his** staat. Hiervoor worden de regressie coëfficiënten uit de file **filter.reg** gebruikt. Na de correctie worden de gecorrigeerde forecasts opgeslagen in de file **filter.cor**. Als de actuele datum **dtm** is, dan geldt:

```

\{\text{dtm}\}\text{tv.fct}    → fct.dat
       $\{\text{dtm}\}\text{ta.his}$    → obs.his
       $\{\text{prevdtm}\}\text{tv.reg}$  → filter.reg

output filter.cor →  $\{\text{dtm}\}\text{ta.cor}$ 

```

Het format van de files $\{\text{dtm}\}\text{tv.fct}$ en $\{\text{prevdtm}\}\text{tv.reg}$ is boven al beschreven. De file $\{\text{dtm}\}\text{ta.his}$ bevat 7 kolommen met elk 24 regels. In de kolommen staat resp.

```
nrstat nrsens yy mm dd hh tobs,
```

in het fortran format (**I4,5(1X,I2),1X,F6.2**). Hier is **nrstat** het nummer van het RWS station waar de waarnemingen vandaan komen, **nrsens** het sensornummer, **yy** het jaar, **mm** de maand, **dd** de dag, **hh** het uur, en **tobs** de waargenomen wegdektemperatuur. De datumtijdgroep, **dtg=yymmddhh**, in deze file begint met **dtg= $\{\text{prevdtm}\}(\text{ta}+1)$** , en eindigt met de waarneming op het aktuele tijdstip **dtg= $\{\text{dtm}\}\text{ta}$** .

De file $\{\text{dtm}\}\text{ta.cor}$ bevat 8 kolommen met elk 24 regels. In de kolommen staat resp.

```
nrstat nrsens yy mm dd hh tfct tfctcorr,
```

in het fortran format (**I4,5(1X,I2),2(1X,F6.2)**). Hier is **nrstat** het nummer van het RWS station waar de waarnemingen vandaan komen, **nrsens** het sensornummer, **yy** het jaar, **mm** de maand, **dd** de dag, **hh** het uur, **tfct** de VAISALA forecasts, en **tfctcorr** de gecorrigeerde forecasts. De datumtijdgroep, **dtg=yymmddhh**, in deze file begint met **dtg= $\{\text{dtm}\}\text{tv}$** .