



Rapportage aan KDC betreffende kwaliteit Statistische modelverwachtingen BZO fasen (KDC LVP project)



Rollende mistbank op Schiphol - 2^e Kerstdag 2010 – foto: Peter J. de Vries

Samenvatting

Verificatie van het statistische TAFguidance model op data van de afgelopen 3 jaar heeft laten zien dat de kwaliteit van de kansverwachtingen hoog is. BZO fase A en B kwamen in de onderzochte periode vaker voor dan het klimatologische langjarige gemiddelde. De betrouwbaarheid van de kansverwachtingen is significant hoger dan de kansverwachtingen in de periode 2004-2007. Ook de analyse van resultaten in geval van een categorische keuze laten een aanzienlijk beter resultaat zien dan de resultaten over de periode voorafgaand aan het KDC LVP project.

Inleiding

In dit memo worden de meest recente verificatieresultaten besproken van de uitkomsten van het statistische TAFguidance model. Dit naar aanleiding van vervolgvragen die gerezen zijn na bespreking van het document “Management samenvatting KDC LVP project” [2] op 25 maart 2011.

Het statistische TAFguidance model is in het kader van het KDC LVP project in 2007-2008 aangepast om de verwachtingen voor LVP condities te verbeteren. De belangrijkste verandering betrof destijds het toevoegen van verwachtingen voor de achtergrondshelderheid, waardoor het mogelijk werd om de kansen voor onderschrijding van zicht drempelwaarden in termen van Runway Visual Range expliciet te berekenen.

Methode

Voor de huidige verificatie is 3 jaar data gebruikt. Het betreft de periode 1 juni 2008 tot 1 juni 2011. Er is gefocust op de prestaties van het model voor de verwachting van de condities in de ochtendpiek. Het verificatietijdstip is dan ook 06 UTC. De geverifieerde verwachtingen zijn uitgegeven kort na 02 UTC, zodat de laatst meegewogen waarneming de waarneming van 02 UTC is. De informatie van het achterliggende atmosfeermodel Hirlam is van de run van de vorige dag 18 UTC, dus 12 uur voor het verificatietijdstip.

Ondanks dat het een periode van 3 jaar betreft is het aantal dagen dat BZO fase A, B of C optrad op het verificatietijdstip, beperkt. BZO fase A of slechter op het tijdstip 06 UTC kwam in de onderzochte periode in circa 5 tot 6% van de dagen voor. Niettemin is dit meer dan het klimatologisch gemiddelde, dat slechts 2,2% bedraagt. Het TAFguidance model levert als uitvoer de kans dat op een bepaald tijdstip (voor deze verificatie 06 UTC) de condities gelijk aan of slechter zijn dan:

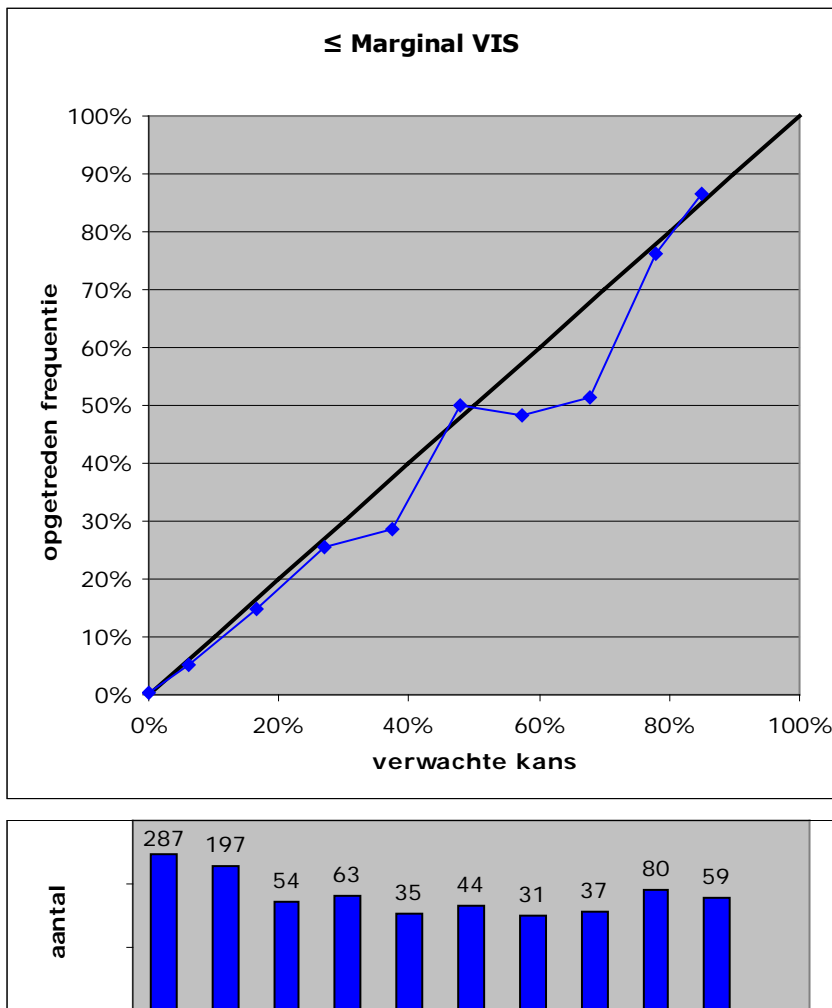
- Marginal
- BZO fase A
- BZO fase B
- BZO fase C

Het voorkomen van BZO fase D is dusdanig zeldzaam dat het, op basis van de huidige beschikbare data, niet mogelijk is om hiervoor een betrouwbare kansverwachting te maken.

Wanneer voor een bepaalde dag een kans van b.v. 40% wordt gegeven dat de condities BZO fase A of slechter zullen zijn, dan is het niet mogelijk deze ene verwachting te verifiëren. Het kan wél of niet zijn uitgekomen, beide uitkomsten zijn mogelijk en niet tegenstrijdig met de kansverwachting van 40%. Met andere woorden: het is dus alleen mogelijk om een groot aantal kansverwachtingen te verifiëren. Wanneer zo'n verwachting van 40% bijvoorbeeld 100 keer wordt uitgegeven, dan zien we graag dat deze in ongeveer 40 gevallen uitkomt en in ongeveer 60 gevallen niet uitkomt. Slechts dan is zo'n kansverwachting betrouwbaar. De kansen in de verwachtingen zijn dan gelijk aan de werkelijk opgetreden frequenties. In dat geval kan de gebruiker hiermee zijn voordeel doen en de kansen toepassen op cost/loss modellen.

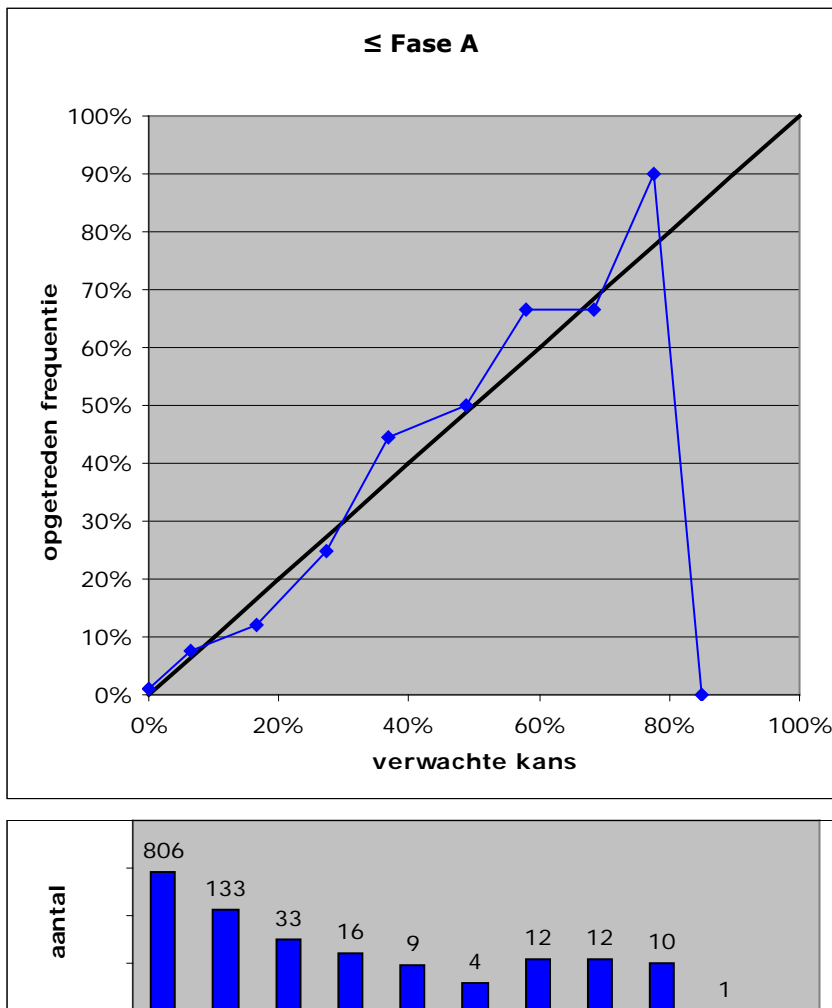
Resultaten

Gezien het bovenstaande heeft het KNMI de betrouwbaarheid van de kansen geanalyseerd die door het statistische model geleverd worden. In onderstaande figuren is langs de horizontale as de verwachte kans uitgezet. Hierbij zijn de verwachtingen geclusterd in segmenten van 10%. Langs de verticale as staat de frequentie van optreden van de uitgegeven kansverwachtingen, oftewel de opgetreden kans. Idealiter liggen de punten dus op de diagonaal in de grafiek: dan zijn de kansen volledig betrouwbaar. De grafieken zijn beschikbaar voor de bovengenoemde 4 condities ($\leq M$, $\leq A$, $\leq B$, $\leq C$). Onder elke grafiek staat het aantal gevallen dat elk punt in de grafiek vertegenwoordigt. Punten in de grafiek die veel gevallen omvatten zeggen uiteraard veel meer dan punten die slechts enkele gevallen betreffen.



Figuur 1. Betrouwbaarheidsdiagram van de kans op Marginal Conditions of slechter (Vis <5000 m of ceiling <1000 ft)

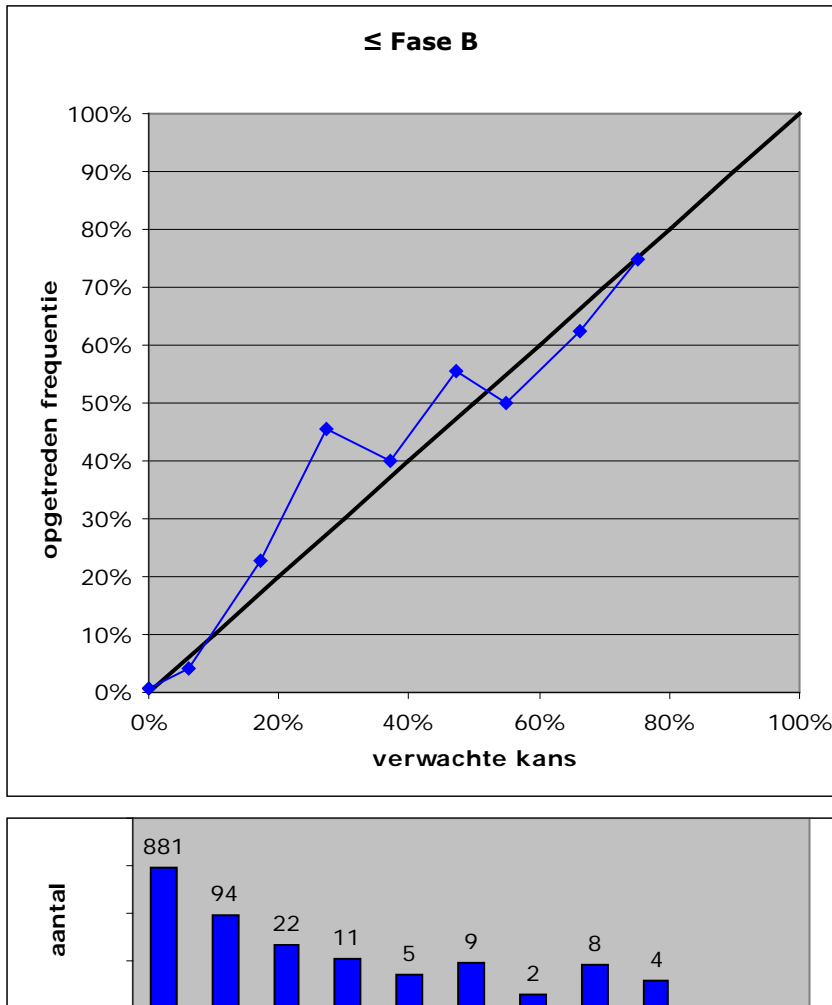
De resultaten in Figuur 1 laten zien dat de kansen die het statistische model berekent, in het algemeen zeer betrouwbaar zijn. Alleen rond 60-70% kans lijkt een lichte overschatting aanwezig te zijn. Verder liggen de resultaten zeer dicht bij de diagonaal, zelfs bij hoge kansen van 80 en 90%, die ook een groot aantal gevallen vertegenwoordigen.



Figuur 2. Betrouwbaarheidsdiagram van de kans op BZO fase A condities of slechter (RVR <1500 m of ceiling <300 ft)

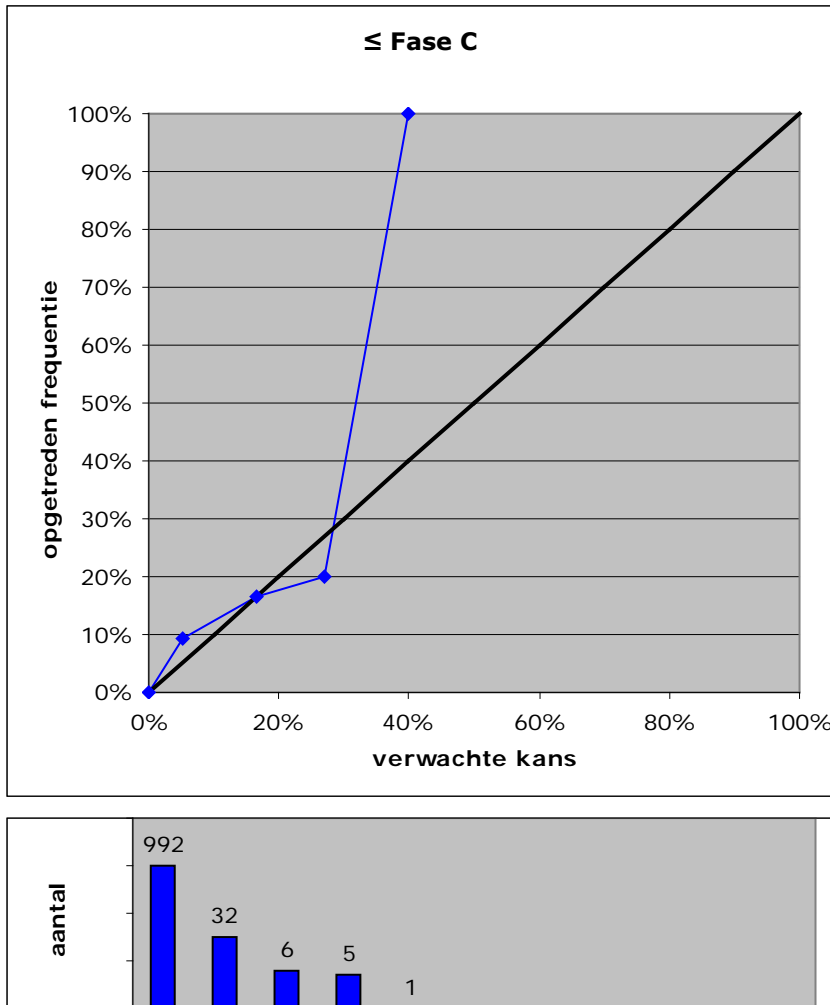
Ook de kansen op fase A of slechter blijken erg betrouwbaar, want ze liggen zeer dicht bij de diagonaal. In één geval is een kans van 85% berekend, en in dat geval is het niet opgetreden. Dit verklaart het ene punt dat ver weg van de diagonaal ligt. Omdat dit slechts 1 geval betreft, heeft het statistisch gezien een zeer beperkte betekenis.

Bovenstaand betrouwbaarheidsdiagram kan niet zonder meer vergeleken worden met Figure 8 van het “final report” uit 2008 [1]. In de betreffende grafiek zijn 8 verschillende uitgiftetijden per dag en 3 verschillende forecasttijden (+3, +6, +9 uur) weergegeven, terwijl we nu alleen hebben gekeken naar de +4 uur verwachting voor het tijdstip 06 UTC.



Figuur 3. Betrouwbaarheidsdiagram van de kans op BZO fase B condities of slechter (RVR <550 m of ceiling <200 ft)

Ook de kansen op fase B of slechter blijken erg betrouwbaar, want ze liggen in het algemeen zeer dicht bij de diagonaal. Het aantal gevallen is beperkt, maar het feit dat de punten dicht bij de diagonaal liggen geeft veel vertrouwen in de berekende kansen.



Figuur 4. Betrouwbaarheidsdiagram van de kans op BZO fase C condities of slechter (RVR <350 m)

Ook de kansen op fase C of slechter blijken betrouwbaar, want ze liggen dicht bij de diagonaal. In één geval werd een kans van 40% berekend, en in dit geval is in werkelijkheid ook fase C opgetreden. Dit punt heeft statistisch geen betekenis, omdat het slechts 1 geval betreft. Merk op dat het in het overgrote deel (96%) van de gevallen terecht een kans van 0% wordt berekend. Fase C of slechter is een zeldzaam fenomeen. In de beschouwde periode van 3 jaar is het slechts 7 keer voorgekomen. Omdat de kans van optreden zo klein is, geeft het statistische model in geen enkel geval een kans van meer dan 40% op het optreden van fase C.

De opgetreden frequentie van de BZO fasen in de onderzochte periode van 3 jaar was hoger dan op grond van de langjarige klimatologische gemiddelden verwacht mocht worden. In Figure 2 van het "final report" uit 2008 [1] is af te lezen dat het gemiddeld aantal dagen met fase A of slechter om 06 UTC, rond 8 dagen per jaar ligt. In deze periode lag dit op bijna 19 dagen per jaar. Ook het aantal dagen met fase B of slechter lag met 12 dagen per jaar veel hoger dan het 30-jaars gemiddelde van 4 tot 5 dagen per jaar. De frequentie van fase C of slechter was wel in overeenstemming met het langjarig gemiddelde van 2 tot 3 dagen per jaar.

Categorische keuze

In de operationele praktijk kan het voorkomen dat een gebruiker vooraf maatregelen wil nemen om de operatie voor te bereiden op een bepaalde BZO fase. In zo'n geval kan een objectieve keuze zijn om uit de volledige kansverdeling één kansdrempel te kiezen en de bijbehorende categorie als verwachting te beschouwen voor de condities op het verwachtingstijdstip. Hoewel men hiermee veel belangrijke informatie negeert, kan het een zinnige keuze zijn als men uit wil gaan van één scenario. Deze strategie is ook besproken in hoofdstuk 6 van het final report uit 2008 [1].

Hieronder staan de contingentietabellen voor twee gevallen. In het eerste geval wordt de categorische keuze gemaakt bij een kansdrempel van 50%, in het tweede geval ligt deze drempel bij 25%.

Verwachting TAFG (p=50)						
Waarneming		C	B	A	GM	
	C	0	3	3	1	7
	B	0	8	8	13	29
	A	0	1	4	15	20
	GM	0	6	5	969	980
	0	18	20	998	1036	

Tabel 1. Contingentie tabel bij categorische keuzedrempel van 50%

Verwachting TAFG (p=25)						
Waarneming		C	B	A	GM	
	C	2	3	1	1	7
	B	3	13	4	9	29
	A	0	3	6	11	20
	GM	1	14	14	951	980
	6	33	25	972	1036	

Tabel 2. Contingentie tabel bij categorische keuzedrempel van 25%

Hits= paars

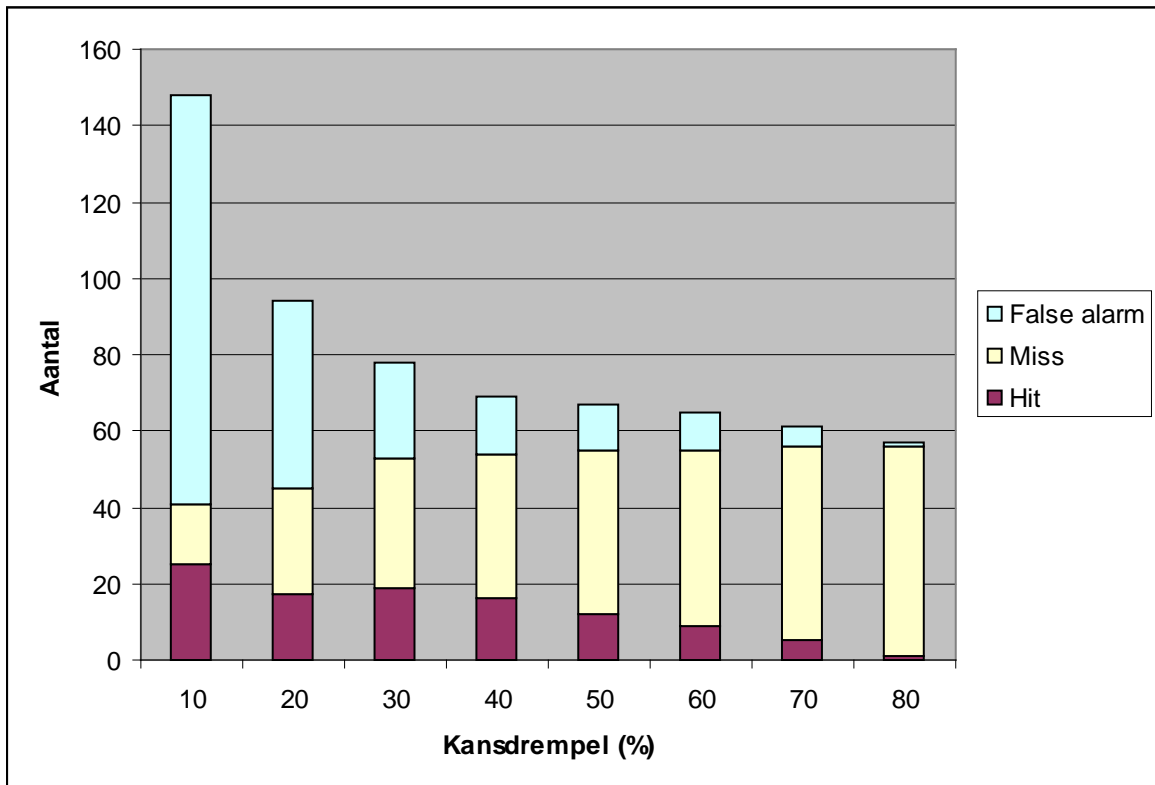
Missers = blauw = waarneming slechter dan verwachting

False alarms = rood = waarneming beter dan verwachting

Uit bovenstaande tabellen blijkt dat verreweg het grootste aantal gevallen zich bevinden in de categorie waarbij Good of Marginal verwacht wordt en dat dit ook daadwerkelijk optreedt. Bij deze kansdrempels ligt het percentage hits in genoemde categorie op 92 tot 94%.

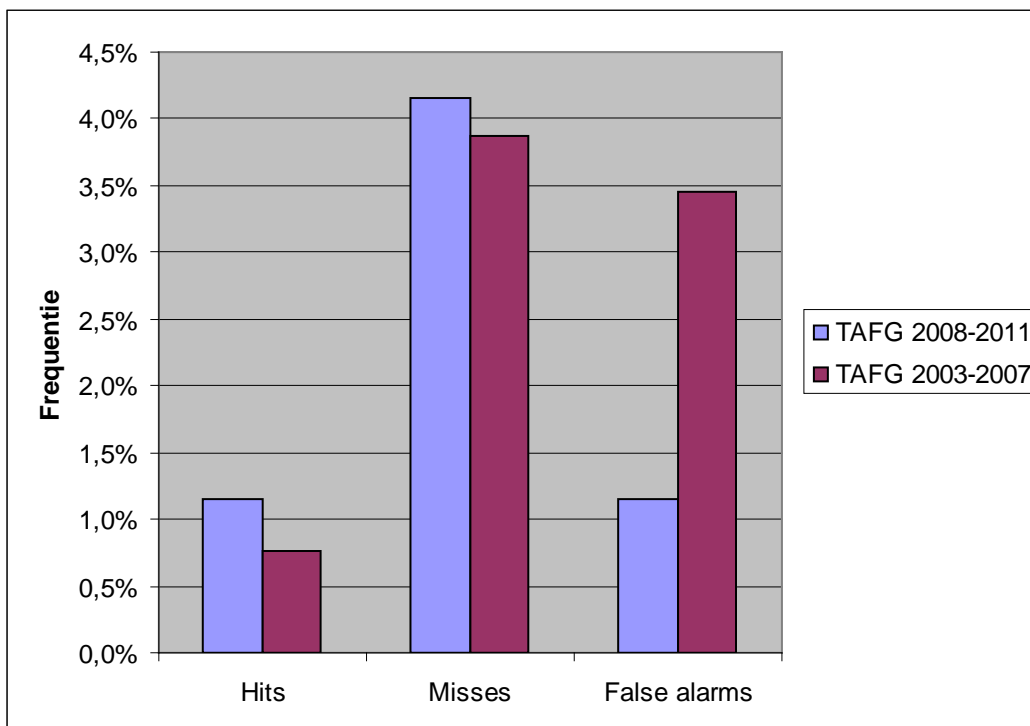
Wanneer we inzoomen op de gevallen die buiten de hierboven genoemde bulk van GM-GM gevallen liggen, dan ziet de verdeling er anders uit. In figuur 5 is weergegeven welke aantallen hits, false alarms en missers opgetreden zouden zijn, indien men een keuze voor een categorie legt bij een kansdrempel van 10 tot en met 80%. Het grote aantal gevallen met Good of Marginal verwacht en opgetreden is hier dus buiten beschouwing gelaten. Indien men de keuze legt bij een laag percentage levert dit veel false alarms op en weinig missers op. Naarmate de kansdrempel hoger wordt gelegd, neemt het aantal false alarms af en het aantal missers toe. Merk op dat de som van hits, false alarms en missers niet constant is, maar steeds lager wordt naarmate de kansdrempel hoger wordt gelegd. Dit komt omdat bij een hogere kansdrempel het steeds vaker voorkomt dat zowel de gekozen categorie als de waarneming in de Good of Marginal categorie ligt, en die is in deze grafiek buiten beschouwing gelaten.

Wanneer men deze keuze voor een enkele categorie zou willen toepassen, is het dus van belang om af te wegen hoeveel missers en hoeveel false alarms acceptabel zijn. Aan de hand van de verhouding van kosten/overlast/nadeel van een “verkeerde keuze” kan men de optimale kansdrempel vaststellen.



Figuur 5. Verdeling tussen aantal hits, false alarms en missers bij diverse categorische kansdrempels. Alleen voor categorieën fase A en slechter.

Tenslotte is ook een vergelijking gemaakt met Figure 12 uit het final report uit 2008 [1]. In onderstaande Figuur 6 is het aantal hits, false alarms en missers vergeleken met de situatie over de periode 2004-2007, dus voordat het verbetertraject werd ingezet.



Figuur 6. Verdeling tussen aantal hits, false alarms en missers bij een kansdrempel van 50%, en vergelijking met de resultaten van voor het KDC LVP project.

In Figuur 6 is te zien dat het aantal missers iets is toegenomen, maar dat het aantal hits significant is toegenomen en dat het aantal false alarms drastisch is gereduceerd ten opzichte van de situatie van voor 2008. Ook in deze grafiek zijn de GM-GM gevallen buiten beschouwing gelaten. Het aantal gevallen is hier uitgedrukt in een percentage om vergelijking mogelijk te maken met de periode 2003-2007. Dat betreft een periode van 4 jaar, terwijl we in deze evaluatie naar een periode van 3 jaar gekeken hebben. Het percentage is ten opzichte van alle gevallen in de tabel met uitzondering van de GM-GM gevallen. Deze figuur is voor een kansdrempel van 50% en heeft duidelijk meer missers dan false alarms. Indien men een andere verhouding wil, kan de keuzedrempel bij een lagere kans worden gelegd. In Figuur 5 is af te lezen dat bij een kansdrempel van b.v. 20% het aantal missers juist kleiner is dan het aantal false alarms.

Literatuur

- [1] "Improved Low visibility and Ceiling Forecasts at Schiphol Airport"; Final Report, KDC-LVP Project Team, July 2008, KDC/2008/0089, KNMI publication 222.
- [2] "Management samenvatting KDC LVP Project"; KNMI memo d.d. 7 maart 2011.