

Naar een conceptuele basis voor omgaan met onzekerheid in stromingsmodellen van Rijkswaterstaat



Naar een conceptuele basis voor omgaan met onzekerheid in stromingsmodellen van Rijkswaterstaat

Auteur(s)

Koen Berends

Ferdinand Diermanse

Naar een conceptuele basis voor omgaan met onzekerheid in stromingsmodellen van Rijkswaterstaat




Opdrachtgever	Rijkswaterstaat Water, Verkeer en Leefomgeving
Contactpersoon	de heer M. Scholten
Referenties	--
Trefwoorden	--

Documentgegevens

Versie	1.0
Datum	21-12-2021
Projectnummer	11206813-018
Document ID	11206813-018-ZWS-0003
Pagina's	39
Classificatie	
Status	definitief

Auteur(s)

	Koen Berends	Jurjen de Jong
	Ferdinand Diermanse	

Doc. Versie	Auteur	Controle	Akkoord	Publicatie
1.0	Koen Berends	Bart van den Hurk	Bianca Peters	
				

Samenvatting

Rijkswaterstaat gebruikt stromingsmodellen in diverse toepassingen en het ondersteunen van beleid, zoals voorspelling van toekomstige waterstanden, vergunningverlening en specialistische onderzoeksvragen. Deze modellen worden ontwikkeld binnen het raamwerk 'kennis voor de primaire processen' (KPP), onderdeel Modelschematisaties.

De modellen worden volgens een vastgelegde procedure gebouwd en afgeregeld. Deze procedure wordt periodiek herzien en omvat naast het stromingsmodellen ook bepalingen voor de te gebruiken databases waarop modellen worden gebaseerd. De nauwkeurigheid van het model wordt binnen deze procedure geoptimaliseerd (door kalibratie) en getest (door validatie).

De geschiktheid van een model voor een toepassing hangt in belangrijke mate af van deze bewezen nauwkeurigheid, die tijdens modelbouw wordt vastgesteld. De nauwkeurigheid kan echter niet eenduidig worden vastgesteld voor alle mogelijke toepassingen. Er zijn bijvoorbeeld toepassingen die inherent ontestbaar zijn doordat metingen die equivalent zijn aan de geproduceerde modeluitvoer onmogelijk zijn om uit te voeren. Gebruikers van modeluitvoer weten doorgaans wel dat modellen geen perfecte voorspellers zijn, en dat de nauwkeurigheid van een model ook niet altijd bekend is. Dit wordt soms 'onzekerheid' genoemd. Er is echter geen duidelijke richtlijn om die onzekerheid van modeluitvoer vast te stellen. Rijkswaterstaat heeft daarom aan Deltares gevraagd om uit te werken waarom en hoe onzekerheid een rol speelt bij haar stromingsmodellen, en daarbij te komen tot een praktische basis om onzekerheid in de praktijk te beoordelen.

In hoofdstuk 2 is ingegaan op de vraag waarom het belangrijk is om de onzekerheid van modeluitkomsten te herkennen en kwantificeren. Aan de hand van voorbeelden uit andere vakgebieden is geschetst dat het negeren van onzekerheid in het verleden heeft geleid tot maatschappelijke discussie, en dat het expliciet benoemen van onzekerheid de waarde van modellen voor de maatschappij kan vergroten. Vervolgens is in hoofdstuk 2.2 een definitie gegeven van wat een 'riviermodel' is en wordt verkend in welke opzichten riviermodellen gevoelig zijn voor onzekerheid.

In hoofdstuk 3.1 is een praktische definitie van onzekerheid voorgesteld. In dit rapport beperken wij ons tot kwantificeerbare of statistische onzekerheid van modeluitvoer. Binnen deze definitie wordt een onderscheid gemaakt tussen modelonzekerheid en voorspelonzekerheid. Modelonzekerheid is gedefinieerd als de onzekerheid in modeluitvoer door onzekerheid van het model, bijvoorbeeld door onzekerheid van modelparameters. Voorspelonzekerheid is het verschil tussen modeluitvoer en metingen. Binnen in de huidige praktijk van riviermodellering is dat analoog is aan het bepalen van modelnauwkeurigheid via validatie.

In hoofdstuk 3.2 zijn deze definities geplaatst in het kader van de toepassingsgebieden van riviermodellen. Hierbij worden toepassingsgebieden onderscheiden langs twee assen van extrapolatie: extrapolatie naar ongeziene modelforcering (zoals extreme hoge- of lage afvoer) en extrapolatie naar een ongeziene staat van het systeem (zoals toepassing na een rivierkundige ingreep). Vervolgens wordt beargumenteerd dat het bepalen van de onzekerheid – analoog aan validatietesten – voor elke toepassing een andere test behoeft.

In hoofdstuk 3.3 is een standaard testformulier (modelbijlage) voorgesteld waarin deze testen per toepassingsgebied worden benoemd.

De doelstelling van een modelbijlage is het inzichtelijk ontsluiten van beschikbare informatie over modelnauwkeurigheid- en onzekerheid voor gebruikers, door het opstellen van een gestandaardiseerd formulier dat inhoudelijk verbonden is aan het terminologisch raamwerk van hoofdstuk 2.

In hoofdstuk 4 is de modelbijlage ingevuld voor een specifieke casus van een Rijkswaterstaat model: het zesde-generatie 2D model voor de Maas. Dit model is relatief goed getest – voor elk toepassingsgebied is er ten minste één test of verkenning beschikbaar. Desalniettemin is vastgesteld dat bij de modelopzet er geen bewuste keuze is gemaakt voor het testen van het model voor specifieke toepassingen. Hierdoor ontbreekt een expliciete onderbouwing van het doel, methode en interpretatie van de testen in het kader van de beoogde toepassing. In hoofdstuk 5.1 is gereflecteerd op de beperkingen van deze studie en is de gekozen definitie van onzekerheid in het kader geplaatst van andere raamwerken. Ook wordt de onzekerheid van verschilanalyses bediscussieerd.

In hoofdstuk 5.2 wordt ten slotte geconcludeerd dat in dit rapport een aanzet is gedaan voor een conceptuele basis om te spreken over onzekerheid die aansluit bij de huidige praktijk van het opzetten en testen van riviermodellen. Er worden vijf aanbevelingen gedaan op basis van de bestudeerde casus:

- 1 Standaardiseer praktijk en documentatie van validatietesten op basis van de in dit rapport voorgestelde conceptuele aanpak.
- 2 Voer extra testen uit om de ‘witte vlekken’ in de Maas casus af te dekken.
- 3 Neem in de naamgevingsconventie een term op voor een overkoepelende modellenfamilie die verschillende schematisaties verbindt.
- 4 Neem het Qf-model dat de bovenstroomse afvoer genereert op in een versiebeheersysteem
- 5 Evalueer deze casus en bouw verder aan een gedragen en gedeelde richtlijn voor het omgaan met onzekerheden.

Inhoud

	Samenvatting	4
1	Inleiding	8
1.1	Aanleiding	8
1.1.1	Stromingsmodellen van Rijkswaterstaat	8
1.1.2	Definitie van een riviermodel	8
1.1.3	De geschiktheid van een model voor een toepassing	10
1.2	Doelstelling	10
1.2.1	Vraagstelling en aanpak	10
1.2.2	Afbakening	10
1.2.2.1	Afbakening van modellen	10
1.2.2.2	Afbakening van onzekerheid	11
2	Onzekerheid van modeluitkomsten en toepassingen	12
2.1	Praktische voorbeelden uit andere vakgebieden	12
2.1.1	Tweede kamer wil uitdrukkelijke vermelding van onzekerheden van RIVM	12
2.1.2	Klokkenluider bij RIVM leidt tot rechtzaak, kamervragen	12
2.1.3	Meerwaarde van modellen & onzekerheid	13
2.2	Redenen waarom onzekerheid een rol speelt bij riviermodellen	14
2.2.1	Scenario's waar onzekerheid een rol speelt bij riviermodellen van Rijkswaterstaat	14
2.2.2	Op welke gebieden zijn riviermodellen gevoelig voor onzekerheid?	14
1.	Wees je bewust van kritische modelaannames	14
2.	Kies de complexiteit die bij het onderwerp en de vraag past	15
3.	Let erop hoe je het model in de markt zet	15
4.	Ken de consequenties	15
5.	Ken de limiet tot waar men op modellen kan vertrouwen	15
2.2.3	Het belang van een helder begrip en definitie van onzekerheden	15
3	Naar onzekerheid in de praktijk	17
3.1	Een praktische definitie van modelonzekerheid	17
3.1.1	Wat bedoelen wij met onzekerheid	17
3.1.2	Wat is de relatie tussen onzekerheid en modelfout?	17
3.1.3	Modelonzekerheid speelt een rol in onbemeten situaties	18
3.2	Onzekerheid van toepassingen	20
3.2.1	De onzekerheidsmatrix	20
3.2.2	Het toetsen van de voorspelonzekerheid	21
3.2.3	Het berekenen van de modelonzekerheid	23
3.2.3.1	Voorwaartse analyse	23
3.2.3.2	Inverse analyse	23
3.2.3.3	Praktische toepassing	24
3.3	Een standaard modelbijlage voor riviermodellen	24
3.3.1	Riviermodellen van Rijkswaterstaat	24
3.3.2	Doelstelling van een modelbijlage	25
3.3.3	Onderdelen van de modelbijlage	25
4	Toepassing op een casus	26

4.1	Het Maas-model	26
4.1.1	Casus & documentatie	26
4.1.2	Modelinformatieformulier	26
4.1.3	Testoverzicht	27
5	Discussie, conclusies en aanbevelingen	29
5.1	Discussie	29
5.1.1	Het gebruik van modellen voor verschilanalyse	29
5.1.2	Bronnen van onzekerheid en relatie tot andere raamwerken	29
5.1.2.1	Verhouding met onzekerheidsbronnen en onzekerheidsraamwerken	29
5.1.2.2	Modelvalidatie, datavalidatie en verificatie	30
5.1.3	Opname van derdenstudies	31
5.2	Conclusie en aanbevelingen	31
	Referenties	34
A	Testen	37

1 Inleiding

1.1 Aanleiding

1.1.1 **Stromingsmodellen van Rijkswaterstaat**

Stromingsmodellen (ook wel: hydrodynamische of hydraulische) modellen zijn wiskundige computermodellen waarmee stroming en waterstanden in rivieren, estuaria en zeeën worden gesimuleerd. Rijkswaterstaat gebruikt stromingsmodellen in diverse toepassingen en voor het ondersteunen van beleid, zoals voorspelling van toekomstige waterstanden, vergunningverlening en specialistische onderzoeksvragen. Deze modellen worden ontwikkeld binnen het raamwerk 'kennis voor de primaire processen' (KPP), onderdeel Modelschematisaties.

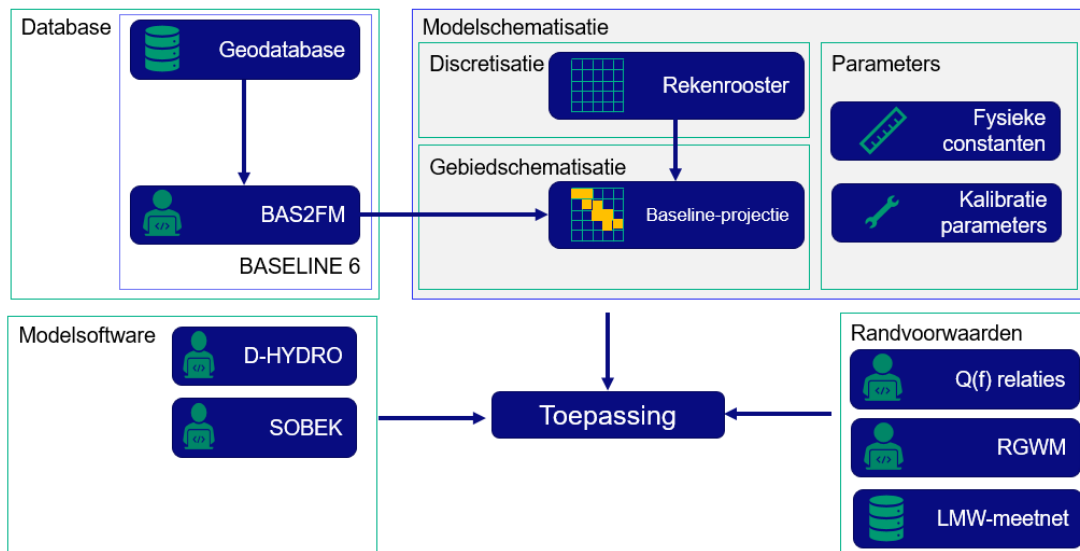
De modellen zijn voor derden vrij beschikbaar (via het webportaal 'Helpdesk Water') om te worden gebruikt voor werk in opdracht van Rijkswaterstaat, onderzoek of andere toepassingen. De aanvrager moet aangeven waarvoor het model gebruikt gaat worden. Rijkswaterstaat kan dan inschatten of het model geschikt is voor het beoogde doel, en waar nodig geeft Rijkswaterstaat de beperkingen van het model aan, levert extra informatie over het model, of schrijft voor welk model gebruikt moet worden voor de beoogde toepassing. Mogelijk past Rijkswaterstaat het model ook aan, zodat het beter geschikt wordt voor een toepassing, bijvoorbeeld door een gebiedsaanpassing door te voeren. In dit rapport beperken wij ons tot riviermodellen.

Riviermodellen worden volgens een bepaalde procedure gebouwd. De reden hiervoor is transparantie en herleidbaarheid. Deze procedure beschrijft bijvoorbeeld hoe een model wordt opgebouwd, op basis van welke data dat gebeurt en hoe een model wordt gekalibreerd. Deze procedure wordt periodiek – ongeveer elke 6 jaar – herzien en aangepast. Elke aanpassing leidt tot een nieuwe generatie modellen. De huidige generatie is de 'zesde'.

1.1.2 **Definitie van een riviermodel**

Een riviermodel is hier gedefinieerd als een digitale (d.w.z. niet fysieke) weergave van een werkelijk systeem die tot doel heeft om de werkelijkheid te beschrijven en de beweging van water en sediment te simuleren. Een 'model' bestaat uit en is afhankelijk van meerdere onderdelen (zie Figuur 1.1 en Tabel 2). Hierdoor kan het moeilijk zijn te bepalen wat wel en wat niet tot het 'model' hoort.

Het is belangrijk om vast te stellen dat een model is gebaseerd op een vastgestelde familie van methodes die in de regel niet verandert binnen een generatie. Zowel de kalibratie als de validatietesten met deze modellen zijn dus afhankelijk van (ook wel: "conditioneel op") deze aanleverende randmethodes. Een significante verandering in deze aanleverende modellen kan leiden tot invalidatie van de testen. Mede hierdoor worden binnen een generatie meestal geen grote veranderingen doorgevoerd.



Figuur 1.1 Een Rijkswaterstaat model bestaat uit verschillende onderdelen.

Tabel 1.1 De verschillende onderdelen en afhankelijkheden van een model. In dit rapport beschouwen we procesmanagers niet.

Onderdeel	Beschrijving	Voorbeelden
Software	Beschrijving van de fysica en de numerieke afhandeling, stromingsmodel, rekenkern, etc.	D-Hydro, SOBEK
Discretisatie	Rekenrooster, resolutie en vorm waarmee de werkelijkheid wordt gerepresenteerd voor toepassing in numerieke modellen	Rekenroosters (2D/3D), netwerken (1D)
Gebiedschematisatie	De vastlegging van gebiedspecifieke informatie naar modelmatige invoer, en die niet door het model zelf wordt aangepast.	Hoogtemodel (2D/3D), Dwarsprofielen, ruwheidsinformatie,
Randvoorwaarden en initialisatie	De forcering van het model, vaak specifiek voor een bepaald <i>scenario</i> dat men doorgerekend heeft, en de initiële condities.	Tijdseries, Q-H relaties
Parameters	Waarden van parameters en instellingen in de modelvergelijkingen, vastgesteld door meting, kalibratie of aanname	Kalibratieparameters, viscositeitsparameters, tijdstapinstellingen
Procesmanager	Een proces manager (script, programma)	OVGK, Delft-FEWS, SGWM
Database	De onderliggende database waarvan een model wordt afgeleid	Baseline

Hoewel het dus moeilijk is te bepalen wat wel en niet tot een model hoort, is het voor dit rapport noodzakelijk om met een definitie te werken. In dit rapport gebruiken we de term 'model' als volgt: een familie van versie-, gebieds- en softwarespecifieke modelschematisaties.

Ter voorbeeld: het "2D Rijntakkenmodel" omvat alle modelschematisaties met de naam **dflow2d-rijn-xxx_6-v1**, waarbij deze naam de *Naamgevings conventies modellen Rijkswaterstaat versie 2.0* volgt. Dit is een model geschikt voor een specifieke software (dflow2d) en een specifiek gebied (de Rijn), met een bepaald versienummer (v1). Onder dit model vallen dus verschillende modelschematisaties, bijvoorbeeld: dflow2d-rijn-j19_6-v1. Deze modelschematisatie beschrijft de Rijn met de actuele gebiedschematisatie van 2019.

1.1.3 De geschiktheid van een model voor een toepassing

De geschiktheid van een model voor een toepassing hangt in belangrijke mate af van de bewezen nauwkeurigheid, dat tijdens modelbouw via validatie wordt vastgesteld. Tijdens validatie worden modelresultaten vergeleken met metingen. Het verschil tussen modeluitkomst en meting is een belangrijke maatstaf voor de nauwkeurigheid.

De nauwkeurigheid kan echter niet eenduidig worden vastgesteld voor alle mogelijke toepassingen. Er zijn bijvoorbeeld toepassingen die inherent ontestbaar zijn doordat metingen die equivalent zijn aan de geproduceerde modeluitvoer onmogelijk zijn om uit te voeren. Een voorbeeld hiervan is het toetsen aan extreme, nog niet eerder opgetreden afvoer, of testen aan een toekomstige situatie (de toekomst kan immers niet bemeten worden).

Gebruikers van modeluitvoer weten doorgaans wel dat modellen geen perfecte voorspellers zijn, en dat de nauwkeurigheid van een model ook niet altijd bekend is. Dit wordt soms 'onzekerheid' genoemd. Er is echter geen duidelijke richtlijn om die onzekerheid van modeluitvoer vast te stellen. Vanuit Rijkswaterstaat is er daarom niet alleen behoefte aan het bepalen (kwantificeren) van onzekerheid, maar ook aan een richtlijn om die te interpreteren in termen van praktische bruikbaarheid van de modelresultaten.

1.2 Doelstelling

1.2.1 Vraagstelling en aanpak

Rijkswaterstaat heeft aan Deltares gevraagd om uit te werken waarom en hoe onzekerheid een rol speelt bij haar stromingsmodellen, en daarbij te komen tot een praktische basis om onzekerheid in de praktijk te beoordelen.

De vraag 'waarom en hoe onzekerheid een rol speelt', behandelen we in hoofdstuk 2 door beschouwing van onzekerheid in andere vakgebieden.

De vraag om te komen tot een praktische basis om onzekerheid in de praktijk te beoordelen, vullen we in de hoofdstukken 3 en 4 in. We stellen ons hierbij niet tot doel om een leidraad te geven voor een systematische verkenning van onzekerheden, maar om te komen tot een gestandaardiseerde documentatie van onzekerheidstesten op basis van een praktisch toepasbare conceptuele basis. In hoofdstuk 3 wordt deze conceptuele basis uiteengezet, en in hoofdstuk 4 wordt deze toegepast op een casus.

1.2.2 Afbakening

Twee begrippen, 'model' en 'onzekerheid', zijn dermate breed en staan dermate centraal in dit rapport dat afbakening van hun betekenis in dit rapport noodzakelijk is.

1.2.2.1 Afbakening van modellen

In dit rapport beperken wij ons tot stromingsmodellen (ook wel *hydrodynamische* of *hydraulische* modellen) van rivieren. Dit laat onverlet dat de hier besproken concepten, zij het met enige vertaling, ook toepasbaar zijn op andere type rekenmodellen (zoals *hydrologische*, *ecohydraulische* of *morfydynamische* modellen) en andere systemen (zoals *estuaria*, *kust- en meersystemen*).

Wat er praktisch wordt verstaan onder "model" is in paragraaf 1.1.2 benoemd.

1.2.2.2 Afbakening van onzekerheid

Het begrip 'onzekerheid' wordt zowel in de omgangstaal als in de wetenschap voor verschillende, vrij uiteenlopende begrippen gebruikt. Zo beschrijft het een emotie (het gevoel van onzekerheid), spraakverwarring (ambigüiteit) en de fundamentele (zoals *Knightsiaanse* en *Heisenbergse*) onzekerheid.

In dit rapport volgen we de definitie van (Walker et al. 2003): "*onzekerheid is elke afwijking van het onbereikbare ideaal van volledige deterministische kennis van het bestudeerde systeem*". We beperken we onszelf binnen deze definitie tot kwantificeerbare onzekerheid (ook wel 'statistische onzekerheid'): onzekerheden waarvan de kansverdelingen van de uitkomsten redelijkerwijs bepaald kunnen worden.

In de literatuur wordt er onderscheid gemaakt tussen kwantificeerbare onzekerheid ('*shallow uncertainty*') en onkwantificeerbare onzekerheid ("*Deep uncertainty*"), die relevant zijn wanneer de kansverdelingen zeer slecht of niet bepaald kunnen worden, of wanneer deze kansverdelingen onkenbaar (*unknowable*) zijn (Stein and Stein 2013). In dit rapport gaan we dus nadrukkelijk niet in op 'deep uncertainty'.

Tot slot richten wij ons op de onzekerheid van modeluitvoer, enerzijds door onzekerheid van het model zélf (bijvoorbeeld door onzekerheid in de waarde van ruwheidsparameters), anderzijds door verschillen tussen model en meting. Andere 'onzekerheden' zoals meetonzekerheid nemen wij niet expliciet mee. Voor de doelstelling van dit rapport stellen we dat deze onzekerheden tot uitdrukkingen komen in de onzekerheid van de modeluitvoer (grote meetonzekerheid kan bijvoorbeeld leiden tot een groter verschil tussen model en meting), en dat ze daarmee een bron zijn van onzekerheid. In paragraaf 3.1 wordt deze definitie verder uitgewerkt.

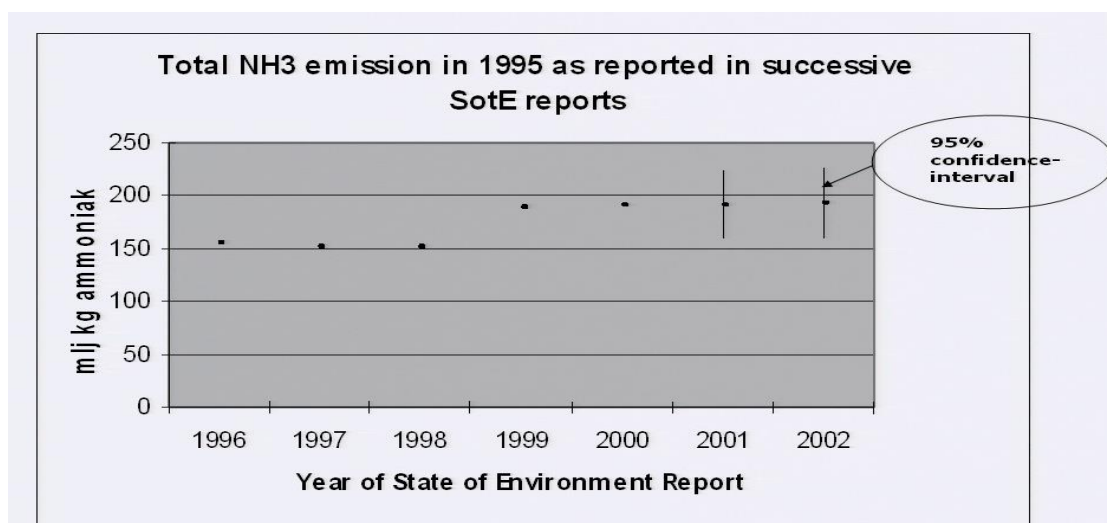
2 Onzekerheid van modeluitkomsten en toepassingen

In dit hoofdstuk wordt aan de hand van voorbeelden uit andere vakgebieden en vakliteratuur geïllustreerd dat het niet voldoende in kaart brengen van de onzekerheid van modelvoorspellingen kan leiden tot kamervragen, rechtszaken en gezichtsverlies. Vervolgens wordt de praktijk van riviermodellen bestudeerd en wordt geconcludeerd dat onzekerheid nog onvoldoende wordt meegenomen.

2.1 Praktische voorbeelden uit andere vakgebieden

2.1.1 Tweede kamer wil uitdrukkelijke vermelding van onzekerheden van RIVM

Figuur 2.1 toont een voorbeeld van het wel/niet expliciet in beeld brengen van onzekerheden. Het betreft een schatting van de totale uitstoot van Ammoniak in het jaar 1995, zoals geschat in verschillende opeenvolgende jaren. In de jaren 1996-2000 is deze uitstoot als een punt weergegeven, dus zonder expliciete weergave van onzekerheden. In het jaar 2001 is men overgestapt op een andere methode van berekenen van de uitstoot en dat resulteerde in een verhoging van de geschatte waarde. Dat was voor velen een onplezierige verrassing die tot kamervragen heeft geleid. De reactie van toenmalig minister Pronk van VROM was dat milieuplanbureau niet altijd alles precies meten, maar geven de berekeningen van het instituut een trend aan. Pronk schreef aan de Tweede Kamer dat het RIVM onzekerheden over metingen en berekeningen wel „uitdrukkelijk” moet vermelden. Dat heeft in 2001 weerslag gekregen in de onzekerheidsband zoals weergegeven in Figuur 2.1.



Figuur 2.1 Ammoniak uitstoot in het jaar 1995, zoals geschat in verschillende opeenvolgende jaren.

2.1.2 Klokkenluider bij RIVM leidt tot rechtszaak, kamervragen

In 1999 zocht een medewerker van het RIVM de publiciteit. Volgens hem was het RIVM in haar rapportages zoals de Milieubalans, die ten grondslag lagen aan beleid, te veel gaan leunen op modelberekeningen. De modeluitkomsten werden te weinig getoetst aan metingen in het veld. De trend die hij zag was: meer modelstudies, minder waarnemingen. Hij stelt dat dit niet per se fout is, maar dat dit wel moet worden aangegeven, maar “In plaats daarvan suggereert het RIVM dat zijn milieucijfers stevig onderbouwd en nauwkeurig zijn”. Hij stelt dat dit ten onrechte is (Trouw 1999).

Deze affaire leidt tot het onmiddellijke ontslag van de klokkenluider, kamervragen en een rechtszaak waarin het RIVM ongelijk krijgt en het ontslag moet terugdraaien (Breedveld 1999). Het RIVM heeft uit deze affaire geleerd en een grote draai gemaakt naar internationaal hoog aangeschreven onzekerheidsstudies (van Asselt et al. 2001; Petersen et al. 2003).

2.1.3 Meerwaarde van modellen & onzekerheid

In een recent Nature-artikel (Saltelli et al. 2020) werd beargumenteerd aan welke voorwaarden een model moet voldoen om van meerwaarde te zijn voor het beantwoorden van maatschappelijke vragen. In dit artikel werden verschillende voorbeelden genoemd, waar modellen niet behulpzaam of zelfs schadelijk waren, doordat men niet voldoende communiceerde of op de hoogte was van de onzekerheid van de modelvoorspellingen.

Zo heeft het Ministerie van Energie van de VS een complex model, bestaande uit 286 gekoppelde sub-modellen, laten bouwen om de schadelijkheid van lozing van radioactief afval te berekenen. Dit model bleek echter waardeloos, toen uit een onzekerheidsanalyse bleek dat één enkele parameter zó onzeker was, dat al het detail dat in het model was gebracht niets meer toevoegde aan het antwoord.

In een meer recent voorbeeld werd het duidelijk dat de onzekerheid van de schatting van het aantal doden als gevolg van COVID-19 van dezelfde orde grootte was als de voorspelling zelf. Hierdoor waren de berekeningen wel nuttig als gedachtenexperiment, maar niet als nauwkeurige voorspelling. Om deze reden moet er een onderscheid gemaakt worden tussen de modeltoepassing. (Klemeš 1986a) onderscheidde korte-termijn-voorspelling, simulatie en lange-termijn-voorspelling. Hij beargumenteerde dat langs deze drie ordes van modelgebruik de complexiteit van het probleem om te spreken over nauwkeurigheid toeneemt, en onze mogelijkheid om goed te testen afneemt. Immers “de mogelijkheid om een verkeerde extrapolatie te corrigeren met metingen komt altijd te laat of komt nooit”.

In het artikel van (Saltelli et al. 2020) werd beargumenteerd dat men op de volgende vijf zaken moet letten bij gebruik van modellen voor vraagstukken uit de samenleving. Deze vijf zaken kunnen als volgt worden samengevat:

- 1 **Wees je bewust van modelaannames**
Doe onzekerheid & gevoeligheidsanalyse
- 2 **Kies de complexiteit doe bij het onderwerp en de vraag past**
Complexiteit is niet altijd een indicator voor kwaliteit
- 3 **Let erop hoe je het model in de markt zet**
All-purpose modellen bestaan niet. Wees heel duidelijk wat een model wel en niet kan.
- 4 **Ken de consequenties**
Onrealistische schattingen van nauwkeurigheid kan leiden tot grote problemen als men daarop vertrouwt
- 5 **Ken de limiet tot waar men op modellen kan vertrouwen**
Een model is niet altijd het juiste gereedschap. Als onzekerheid te groot is (zie ook diepe onzekerheid), moet men andere technieken gebruiken.

2.2 Redenen waarom onzekerheid een rol speelt bij riviermodellen

2.2.1 Scenario's waar onzekerheid een rol speelt bij riviermodellen van Rijkswaterstaat

Ook bij toepassingen van riviermodellen zijn situaties te bedenken waarbij de onzekerheid rondom een modelvoorspelling belangrijk kan zijn.

- Als mensen gedwongen moeten verhuizen om plaats te maken voor een rivierverruimende maatregel, bieden modelvoorspellingen van het effect van die maatregel op maatgevende waterstanden in de orde van centimeters dan wel voldoende bewijs voor de noodzaak en effectiviteit van die maatregel? Zou dit argument winnen in een rechtszaak, of in de publieke opinie?
- Als de dijken langs de rivier ontworpen zijn op waterstanden die optreden bij een afvoer van $4000 \text{ m}^3\text{s}^{-1}$, is evacuatie bij lagere afvoeren dan absoluut niet nodig?
- Is het vertrouwen in de manier waarop modellen de ruwheid van vegetatie in de uiterwaarden modelleren dusdanig hoog, dat modelvoorspellingen voldoende bewijs leveren voor het wel of niet beschermen van lokale flora en fauna?

Het is altijd lastig om dergelijke uitspraken te doen. In de praktijk kan vaak worden teruggevallen op experts of strikte beslisprotocollen waarbij modeluitkomsten als leidend worden beschouwd. Desalniettemin bestaat er een reëel risico in de perceptie dat men zich 'verschuilt achter modelberekeningen'. Hoewel al deze vragen tot op zekere hoogte afgevangen (kunnen) worden door procedures, leert het verleden (zie paragraaf 2.1.2) dat het vertrouwen in een instituut ernstig ondergraven kan worden als onzekerheid niet voldoende in kaart wordt gebracht en wordt gebruikt bij de interpretatie van (de representativiteit van) modeluitkomsten.

Zelfs als het bij belanghebbenden wel helder voor ogen staat dat modelvoorspellingen met onzekerheden behept zijn, is het nog altijd relevant om een orde van grootte van de mate van onzekerheid te kunnen vermelden. Ze kunnen deze immers besluiten of ze op basis van de modelresultaten verstrekkende beslissingen durven te baseren. Daarbij zouden ze dan bij voorkeur gevoed moeten worden door gevoeligheidsanalyses die laten zien wat de gevolgen kunnen zijn van eventuele betekenisvolle afwijkingen van de verwachte uitkomst (zoals hogere waterstanden dan voorspeld tijdens een hoogwater, of juist lagere waterstanden tijdens een laagwaterperiode).

2.2.2 Op welke gebieden zijn riviermodellen gevoelig voor onzekerheid?

Aan de hand van de vijf overdenkingen van (Saltelli et al. 2020) reflecteren we in hoeverre de modellen van Rijkswaterstaat vatbaar zijn voor onzekerheid.

1. Wees je bewust van kritische modelaannames

De relevante vraag is hier: zijn er kritische modelaannames die zowel gevoelig (verschillende waarden leiden tot significant andere modeluitvoer) als onzeker (de waarden zijn niet met voldoende nauwkeurigheid vast te stellen) zijn?

Er is zowel kwantitatief als kwalitatief onderzoek gedaan naar enkele kritische aannames bij Nederlandse riviermodellen. Deze kritische aannames omvatten onder meer vegetatieweerstandformules (J J Warmink, Booij, et al. 2013), ecotopenkaarten (Straatsma et al. 2013), bedruwheidsparameters (J J Warmink 2014), kalibratiemethodes (Domhof et al. 2018), kunstwerkformuleringen (Mosselman and Le 2016), verborgen aannames in modeltoepassingen zoals geen onzekerheid bij verschilanalyses (Koen D Berends et al. 2018; Mosselman and Le 2016), niet-stationariteit (K. Berends 2019) en randvoorwaarden (Twijnstra et al. 2020). Uit deze onderzoeken blijkt dat er meerdere aannames die zowel gevoelig als onzeker zijn.

2. Kies de complexiteit die bij het onderwerp en de vraag past

De relevante vraag is of de complexiteit van riviermodellen (en de relatieve complexiteit van de onderliggende modules) in verhouding staat tot de toepassing. Deze vraag is veel moeilijker te beantwoorden, hoewel er wel degelijk een trend naar het vergroten van de complexiteit. Denk hierbij bijvoorbeeld aan de resolutie van ecotopenkaarten, waarbij de vraag gesteld kan worden in hoeverre het detail in deze kaarten meerwaarde heeft, gelet op onzekerheid van de vegetatiemodellen die ze bedienen.

Een andere vorm van complexiteit is het koppelen van meerdere modellen, waarbij de 'onzekerheidsketen' niet voldoende in beeld is. Denk hierbij bijvoorbeeld aan modellen die randvoorwaarden genereren voor andere modellen. Een concreet voorbeeld is onzekerheid in de afvoermeting op basis van 'Q-H' relaties (Twijnstra et al. 2020). In het algemeen kan men stellen dat er een balans moet worden gezocht tussen de complexiteit van alle met elkaar samenhangende onderdelen.

3. Let erop hoe je het model in de markt zet

Rijkswaterstaat stelt haar modellen beschikbaar via Helpdesk Water. Daarmee kunnen de modellen gebruikt worden in de praktijk voor adviesprojecten en onderzoek. Rijkswaterstaat beoordeelt of een model geschikt is voor de toepassing, en past het model aan als dat nodig is, en geeft bij de uitlevering van een model informatie mee over het beoogde gebruik. Dat is belangrijk om te voorkomen dat er onrealistische verwachtingen ontstaan (bijvoorbeeld die van een *all-purpose* model). Desalniettemin geven modelbeheerders van Rijkswaterstaat aan onvoldoende richtlijnen te hebben te aanzien van onzekerheid. Om gebruikers beter te kunnen bedienen, is een dergelijke richtlijn dus nodig.

4 Ken de consequenties

Door de hoge eisen die worden gesteld aan de kalibratie (en validatie) van modellen kan onbedoeld een beeld ontstaan van modeluitkomsten die op enkele centimeters (of zelfs millimeters) nauwkeurig zijn. Bijvoorbeeld: voorspellingen van rivieringrepen worden zeer nauwkeuriger gecommuniceerd (d.w.z., zonder enige onzekerheid). De optiek van onredelijk groot vertrouwen in modellen kan schadelijk zijn. Het is daarom van meerwaarde om goed over onzekerheid te communiceren.

5 Ken de limiet tot waar men op modellen kan vertrouwen

(Klemeš 1986a) definieerde drie ordes van modeltoepassingen. De derde orde, lange-termijn-voorspellingen, was het meest uitdagend. Voor voorspellingen over de verre toekomst, of over zeer extreme omstandigheden, kan men zich afvragen of de onzekerheden zo groot worden dat men er niet meer op kan vertrouwen. Bijvoorbeeld doordat er onbekende veranderingen of omstandigheden kunnen optreden. Het is belangrijk om te weten dat modellen een beperkt toepassingsgebied hebben. Daarbuiten kan men beter praten over onzekerheid in termen als 'diepe onzekerheid', 'adaptiepaden' en 'veerkrachtige rivieren'. Hoewel dit een belangrijk onderwerp is, beperken wij ons in dit rapport tot modelonzekerheid, en gaan wij niet verder in op 'diepe onzekerheid'.

2.2.3 Het belang van een helder begrip en definitie van onzekerheden

Op 12 december 2019 is een sessie georganiseerd met D-Hydro experts om te komen tot een schatting van onzekerheden in met D-Hydro berekende waterstanden voor de Rijntakken.

Op basis van deze sessie zijn onder andere de volgende conclusies getrokken:

- Er zijn veel verschillende typen onzekerheden. Bovendien zijn er verschillende modeltoepassingen met onderling verschillende model- en voorspelonzekerheden. Eén generiek antwoord op de vraag hoe groot de onzekerheden zijn is daarom niet te geven. Om het toch praktisch te houden is het aan te bevelen om de diverse modeltoepassingen te clusteren, voor elke cluster een orde grootte van de onzekerheid te schatten en vast te stellen of er een vereiste limiet is waarbij een 'fit for purpose' statement kan worden afgegeven.
- Een ander gevolg van het feit dat er verschillende typen onzekerheden zijn, is dat ondervraagde experts een verschillend beeld kunnen hebben van wat onzekerheden zijn. Het gezamenlijk beschouwen van schattingen van meerdere experts kan daardoor een kwestie worden van appels met peren vergelijken. Het is daarom aan te raden om aan het begin van een "expert elicitation"¹ voldoende tijd in te ruimen om te komen tot een gedeelde definitie van onzekerheid.
- Uit de schattingen van experts blijkt dat er over het algemeen redelijke consensus is over de grootte van modelonzekerheden voor gemiddelde condities. Tegelijkertijd zijn er soms ook substantiële verschillen tussen de schattingen van de verschillende experts voor meer extreme condities. Bijvoorbeeld: tussen de grootste en kleinste breedte van de geschatte betrouwbaarheidsintervallen van de onzekerheid in de waterstand zat in meerdere gevallen zelfs een factor 3 verschil. Dergelijke verschillen kunnen in termen van advies en beleid potentieel grote gevolgen hebben; denk bijvoorbeeld aan de invloed van onzekerheden op maatgevende waterstanden die in het kader van WBI zijn afgeleid. Onderlinge verschillen kunnen mogelijk deels verklaard worden door een verschil in perceptie van wat de modelonzekerheid precies behelst (zie vorige bullets).

Een concrete manier om de verschillende typen onzekerheid in beeld te brengen wordt gegeven in het volgende hoofdstuk.

¹ Dit is een geformaliseerde methode waarin schattingen van een groep experts gebundeld worden tot een gezamenlijke uitkomst, inclusief onzekerheidsbanden.

3 Naar onzekerheid in de praktijk

In het vorige hoofdstuk is beschreven waarom het belangrijk is om onzekerheid van modeluitkomsten te beschrijven. In dit hoofdstuk wordt een praktische definitie van onzekerheid van uitkomsten van riviermodellen van Rijkswaterstaat gegeven. Vervolgens wordt gereflecteerd op de huidige praktijk, waarna wordt geconcludeerd dat er een behoefte is om betere uitleg te geven van de beschikbare informatie over nauwkeurigheid en onzekerheid.

Vervolgens wordt een voorstel gedaan voor een standaardbijlage die bij oplevering van een model wordt meegeleverd. Het doel van zo'n formulier is om de geteste nauwkeurigheid te documenteren en een toelichting te geven voor modelgebruik bij ongeteste toepassingen. Dit wordt gedaan door in duidelijke termen te communiceren wat de 'onzekerheid' van een model is, en hoe hiermee kan worden omgegaan voor de diverse toepassingen.

3.1 Een praktische definitie van modelonzekerheid

3.1.1 Wat bedoelen wij met onzekerheid

In paragraaf 0 is gesteld dat we ons in dit rapport beperken tot de onzekerheid in modeluitvoer, enerzijds door onzekerheid in het model zélf en anderzijds door verschillen tussen model en meting. Deze definitie werken we in deze paragraaf verder uit.

De (on)-nauwkeurigheid van modellen wordt ook wel 'onzekerheid' genoemd, waarmee wordt bedoeld dat modelvoorspellingen kunnen afwijken van de werkelijkheid, en dat de grootte van deze afwijking niet of slechts ten dele bekend is. In dit document noemen we de afwijking tussen model en werkelijkheid **totale onzekerheid**. In dit hoofdstuk wordt deze totale onzekerheid uitgesplitst in twee componenten²: **voorspelonzekerheid** en **modelonzekerheid**. We zullen uitleggen dat voorspelonzekerheid datgene is wat traditioneel (d.w.z., in modelbouwrapporten zoals (Becker 2012)) wordt gezien als modelnauwkeurigheid. Daarna zullen we uitleggen waarom modelonzekerheid belangrijk is bij extrapolatie naar onbemeten omstandigheden, en zelfs de totale onzekerheid kan domineren.

3.1.2 Wat is de relatie tussen onzekerheid en modelfout?

Het verschil tussen meting en model is een maat voor de onzekerheid van een modeluitkomst. Dit noemen we ook wel de modelfout. In formulevorm:

$$\text{meting} = \text{modeluitkomst} + \text{fout}$$

Als het model goed is, heeft de fout een eenvoudig stochastisch (willekeurig) karakter. De fout kan dan doorgaans afdoende beschreven worden met de bias en standaardafwijking van een normale kansverdeling. Als het model minder goed is, heeft de fout een ingewikkelder karakter. De fout kan bijvoorbeeld afhankelijk zijn van de afvoer. We spreken daarom ook wel over een 'foutmodel' (*EN: error model*). Dit model wordt door statistici ook wel de **voorspelonzekerheid** genoemd. We kunnen dan dus ook schrijven:

$$\text{meting} = \text{modeluitkomst} + \text{voorspelonzekerheid}$$

² We attenderen de lezer erop dat er in de literatuur een groot scala aan onderverdelingen en benamingen bestaat (Skinner, Rocks, and Pollard 2014). Er is niet één juist onderscheid, maar veel *frames* die, afhankelijk van de toepassing, meer of minder inzicht kunnen bieden. Om te voorkomen dat de lezer verdwaalt in terminologie, kiezen we hier voor een praktisch onderscheid op hoog niveau, zonder de onderliggende oorzaken te benoemen.

Ondanks de naam kan voorspelonzekerheid prima berekend worden. Zoals uit de formule blijkt, zijn daarvoor enkel metingen en een model nodig. In de praktijk wordt het bepalen van de voorspelonzekerheid validatie, verificatie of testen genoemd. Als het model, het systeem en wijze van meten niet erg veranderen³ kan worden aangenomen dat de voorspelonzekerheid het residu tussen modeluitkomst en meting goed beschrijft. Dat is echter niet altijd het geval.

3.1.3 Modelonzekerheid speelt een rol in onbemeten situaties

De keerzijde van bovenstaande laat zich raden: voor veel toepassingen zijn metingen niet of heel beperkt beschikbaar. Denk bijvoorbeeld aan extreme afvoeren, stroomsnelheden in uiterwaarden, toekomstverwachtingen of ingrepeffectstudies. Het kan zijn dat een model het prima doet onder gemeten omstandigheden, maar onder andere omstandigheden systematisch afwijkt.

In dat soort gevallen kunnen we er niet zomaar erop vertrouwen dat de formule **meting = modeluitkomst + voorspelonzekerheid** nog klopt. Als het model of het gemodelleerde systeem immers afwijkt, wijkt de voorspelonzekerheid mee af.

Een manier om hiermee om te gaan, is het introduceren van **modelonzekerheid**. Wij nemen aan dat het model in principe in staat moet zijn om de onbemeten situatie goed te simuleren, maar we weten niet welke instellingen (waaronder in principe alles in Tabel 1.1 kan vallen) we daarvoor moeten gebruiken. Door variatie in modelinstellingen te introduceren, zal er ook variatie komen in de modeluitvoer: modelonzekerheid. Doordat het model *zelf* 'onzeker wordt', veranderd de formule ook:

$$\text{meting} = \text{modelonzekerheid} + \text{voorspelonzekerheid}$$

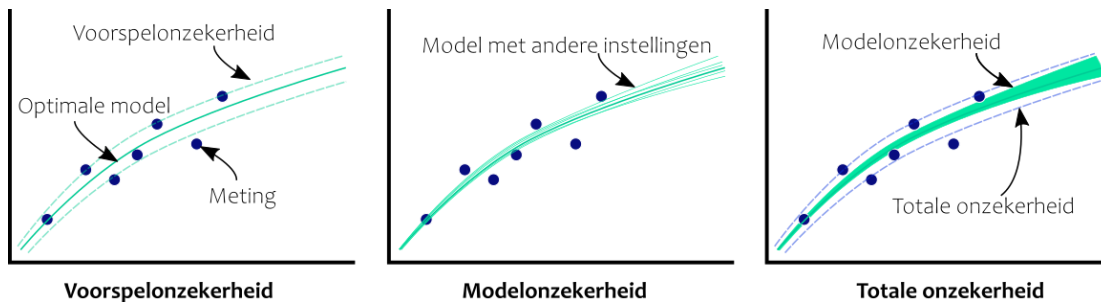
Als we deze theorie toepassen op een relatief eenvoudige relatie (bijvoorbeeld een Q-H relatie), kan dit eruit zien als Figuur 3.1. We plaatsen twee kanttekeningen bij de manier hoe modelonzekerheid in dit figuur is weergegeven. Ten eerste definiëren we modelonzekerheid hier als het *resultaat* van variatie in de instellingen, niet als de *oorzaak*⁴. Ten tweede is de modelonzekerheid in dit voorbeeld *geconditioneerd* op beschikbare meetdata. Dit is echter niet noodzakelijk⁵.

Tot slot merken we op dat in de praktijk vaak deterministisch gekalibreerd wordt, waardoor alle instellingen uiteindelijk worden teruggebracht naar een enkele waarde, waarbij geprobeerd wordt het verschil tussen model en meting te minimaliseren. Daarom zorgt deterministische kalibratie ervoor dat de modelonzekerheid nul wordt, en de voorspelonzekerheid zo klein mogelijk. Een modelonzekerheid die niet gelijk is aan nul, veronderstelt dus dat men niet met slechts één gekalibreerd model rekt.

³ Praktisch gezien betekent dat hier: toepassing onder gelijke omstandigheden. Daarover later (hoofdstuk 3) meer.

⁴ De oorzaak kan bijvoorbeeld zijn: variatie van de ruwheidsparameter van de rivierbodem. In de literatuur worden zowel oorzaken als het gevolg ervan aangeduid als een vorm van modelonzekerheid.

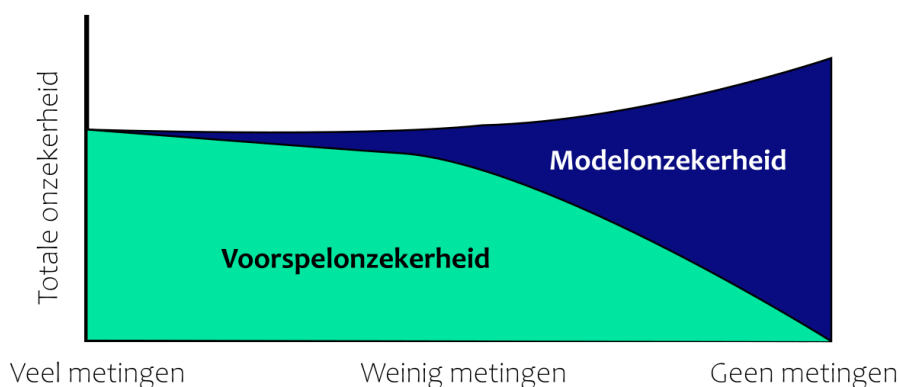
⁵ De instellingen die gevarieerd worden, worden gevarieerd binnen bepaalde grenzen. Waar die grenzen gesteld worden, moet verantwoord worden. Dat kan bijvoorbeeld op basis van directe metingen van parameters (J J Warmink, Booij, et al. 2013). Onder bepaalde omstandigheden kunnen instellingen ook worden afgeleid uit metingen (Koen D. Berends et al. 2020).



Figuur 3.1 Het verschil tussen model-, voorspel- en totale onzekerheid. In dit voorbeeld van kunnen de assen worden geïnterpreteerd als afvoer op de horizontale as en waterstand op de verticale as.

Hoewel Figuur 3.1 is toegepast op een relatief eenvoudig model, blijven de principes gelijk bij toepassing op een riviermodel. Een paar belangrijke observaties die we hieruit leren:

- Wanneer metingen beschikbaar zijn is de modelonzekerheid erg klein, omdat ieder alternatief model ook de metingen volgt doordat ze daarop gekalibreerd zijn. Daarom voegt het berekenen van modelonzekerheid hier niet zoveel toe⁶.
- Hoe minder metingen, hoe belangrijker het aandeel van modelonzekerheid in de totale onzekerheid wordt. Bij grote afstand, domineert modelonzekerheid de totale onzekerheid (Figuur 3.2).



Figuur 3.2 Het aandeel van voorspel- en modelonzekerheid in de totale onzekerheid.

⁶ Dit kan tegen intuïtief zijn omdat ook in het bemeten bereik het model niet perfect is, en omdat metingen ook onzeker zijn. Hoe kan de modelonzekerheid dan naar nul gaan bij heel veel metingen? Het onderliggende proces is wat in de wiskunde bekend is als de centrale limietstelling. Bij weinig metingen is onze schatting van het gemiddelde van een gemeten grootheid (bijvoorbeeld de lengte van rivieringenieurs) onzeker. De mate van onzekerheid kan zelfs worden bepaald door middel van de standaardfout. Maar hoe meer metingen, hoe nauwkeurig onze schatting van het gemiddelde is. In het meest extreme geval waarin we de hele populatie meten, weten we precies wat het gemiddelde is. Op deze manier gaat de onzekerheid van het gemiddelde naar nul. Als analoog kan dit ook zo worden toegepast op modellen, maar in plaats van het gemiddelde gaan we op zoek naar de waarde van parameters (zoals ruwheid). Hoe meer metingen, hoe nauwer de bandbreedte van de parameterwaarden zijn die leiden tot een goede overeenstemming met de meetgegevens.

3.2 Onzekerheid van toepassingen

3.2.1 De onzekerheidsmatrix

Riviermodellen worden gebruikt voor verschillende toepassingen, zoals het (real-time) voorspellen van waterstanden in de nabije toekomst, het simuleren van het effect van (geplande) veranderingen in het systeem, of het bestuderen van het systeemgedrag. Voor sommige van deze toepassingen zijn metingen beschikbaar; voor andere niet. Het hangt dus van de toepassing af, wat voor soort onzekerheid een rol speelt. Om hier handen en voeten aan te geven, adopteren we de onzekerheidsmatrix (K. Berends 2019), waarin de verschillende toepassingen kunnen worden getekend (Figuur 3.3).

De matrix bevat twee assen. Op de horizontale as zit de **modelschematisatie**, waaronder voor riviermodellen alles van Tabel 1.1 valt met uitzondering van de randvoorwaarden (want die vallen onder modelforcering, zie volgende tekstblok). Binnen deze as zijn er twee categorieën: de schematisatie van een (rivier)systeem uit het verleden of één van de toekomst. Een 'toekomstige' schematisatie hoeft niet per se de toekomst te beschrijven, maar kan ook een mogelijk scenario voorstellen.

Op de verticale as ligt **modelforcing** (EN: *model forcing*). Ook hier zijn er twee categorieën: forcering (bijvoorbeeld: afvoer bij Lobith) waarvoor metingen beschikbaar zijn, en forcering waar deze niet beschikbaar zijn. Onder modelforcing vallen bijvoorbeeld de randvoorwaarden.

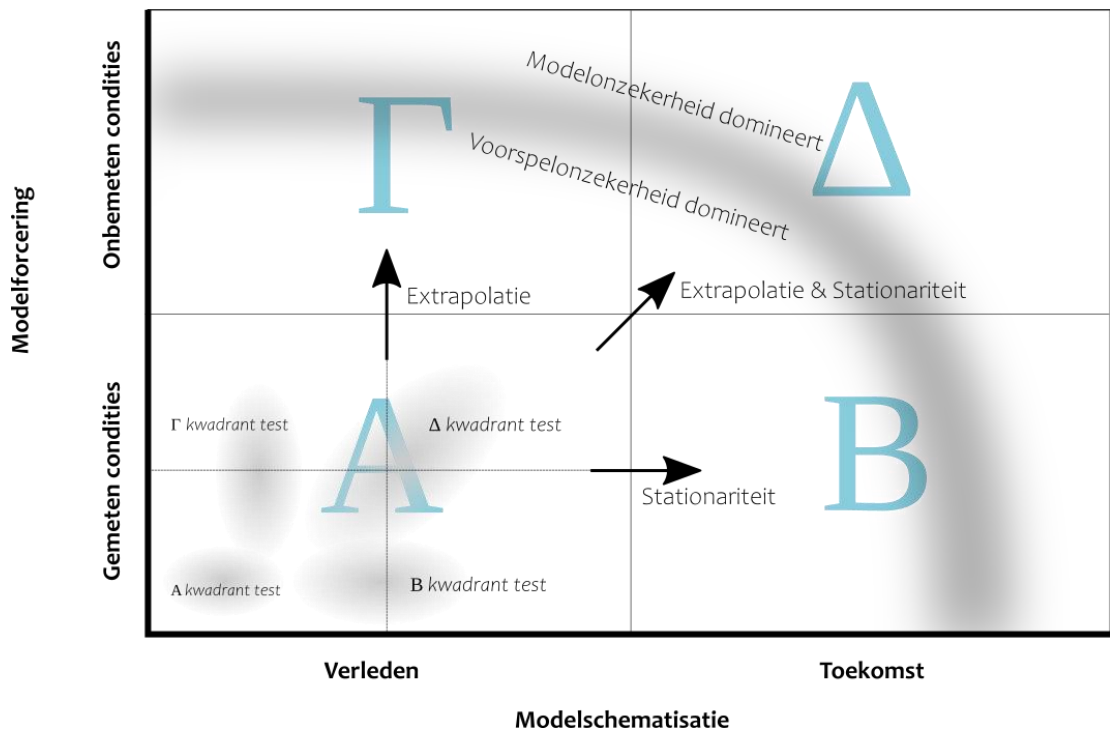
Beide assen leiden dus tot vier kwadranten, weergegeven met de Griekse symbolen voor alfa, beta, gamma en delta. Het is belangrijk om hier het onderscheid te maken, dat elke vorm van toetsing (d.w.z. – vaststellen van de voorspelonzekerheid) *per definitie* in kwadrant A plaatsvindt – dat is het enige kwadrant waar metingen beschikbaar zijn om aan te toetsen.

Toepassingen van de modellen vinden daarentegen vaak plaats in de andere drie kwadranten. Zo is het simuleren van waterstand bij extreme afvoer doorgaans een vorm van extrapolatie (Γ -kwadrant).

Het voorspellen van de stroomsnelheden na een ingreep in de rivier om te zien hoe het de scheepvaart beïnvloedt, bijvoorbeeld in het kader van vergunningverlening, is een berekening van een niet-stationair systeem (B-kwadrant). Niet-stationariteit slaat hier op verschillende maar samenhangende fenomenen (Beven 2016): (1) niet-stationariteit van het riviersysteem door menselijke ingreep – de rivier verandert fysiek – dat mogelijk kan leiden tot (2) niet-stationariteit van het verschil tussen model en meting. Ook (3) niet-stationariteit van randvoorwaarden, bijvoorbeeld door bodemerosie (Koen D. Berends et al. 2021), kan leiden tot niet-stationariteit tussen model en meting.

Een voorbeeld van een combinatie van extrapolatie en (niet-)stationariteit wordt gevonden bij berekeningen van waterstanden na rivierverruimingsmaatregelen bij maatgevende afvoeren die zich nog nooit hebben voorgedaan (Δ -kwadrant). Een speciaal geval is wanneer er wordt gekeken naar het relatieve effect, waarbij wordt gekeken naar het verschil tussen twee berekeningen in soms andere kwadranten. Hierop komen we in de discussie terug (paragraaf 5.1.1).

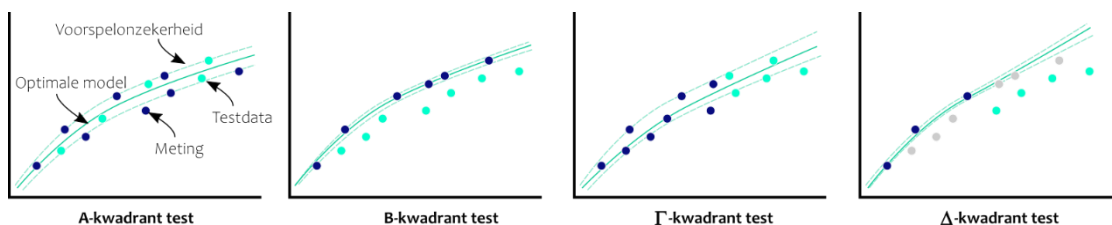
Voor al deze toepassingen kan worden aangenomen dat er een overgangsgebied is waar modelonzekerheid de totale onzekerheid gaat domineren, analoog aan Figuur 3.2. Hoe verder de toepassing afstaat van het meetbare, hoe groter de kans dat modelonzekerheid domineert.



Figuur 3.3 De onzekerheidsmatrix, naar (K. Berends 2019)

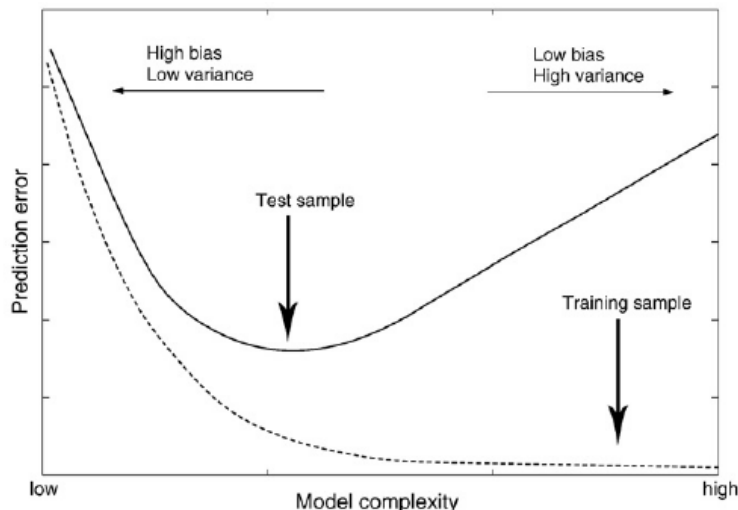
3.2.2 Het toetsen van de voorspelonzekerheid

Modellen kunnen enkel getoetst worden in het A-kwadrant, omdat enkel daar gegevens beschikbaar zijn. Er zijn echter wel degelijk methodes om te testen wat de voorspelonzekerheid is bij toepassing in de overige kwadranten, door binnen kwadrant A van figuur 2.4 de beschikbare data te splitsen in een set waarop het model wordt geconditioneerd (gekalibreerd) en een set waarop het model wordt getoetst. De wijze van het splitsen van de data bepaalt voor welke toepassing het model getest wordt. Deze testen zijn schematisch weergegeven in Figuur 3.4.



Figuur 3.4 Door de data op te splitsen kan getest worden voor verschillende toepassingen. Deze voorbeelddataset bestaat uit metingen uit twee verschillende perioden. In de B-kwadrant test wordt bijvoorbeeld gekalibreerd op de ene periode, en getest op de andere periode. In dit voorbeeld van kunnen de assen worden geïnterpreteerd als afvoer op de horizontale as en waterstand op de verticale as.

Het doel van een **A-kwadranttest** is om te testen op “overfitting”. Hiermee kan bijvoorbeeld getest worden of het model niet te complex is voor de beschikbare data (Figuur 3.5). Deze testen worden uitgevoerd op een zo homogeen mogelijke dataset, waardoor achteruitgang in modelnauwkeurigheid door extrapolatie wordt uitgesloten.



Figuur 3.5 Met een A-kwadrant test kan worden getest wanneer de nauwkeurigheid in validatie ('test sample') afneemt met toenemende nauwkeurigheid in kalibratie ('training sample')

Deze test wordt met name gebruikt in *data science*, voor bijvoorbeeld multivariate regressiemodellen of neurale netwerken. Hierbij verhoogt het aantal variabelen of *hidden nodes* de complexiteit en nauwkeurigheid tijdens kalibratie, maar kan een te grote complexiteit tot verlies van nauwkeurigheid in validatie leiden. Als dit gebeurt is er sprake van overfitting. In eerder onderzoek (Domhof et al. 2018) is getoetst of dit fenomeen ook optreedt bij een riviermodel van Rijkswaterstaat gebleken. Uit dit onderzoek bleek dat voor de manier waarop deze riviermodellen gekalibreerd worden overfitting waarschijnlijk geen gevaar is. Deze manier van testen is ook wel bekend als (*kruis*)validatie (EN: *cross-validation of split-sample-test*).

Het doel van een **B-kwadranttest** is om te toetsen of de voorspelonzekerheid gelijk blijft wanneer het onderliggende systeem verandert. Een eerste aanzet voor zo'n test werd gegeven door (Thirel, Andréassian, and Perrin 2015). Een andere manier om deze test te omschrijven is om te testen of de modelfout stationair is, wanneer het systeem niet-stationair is. Over het algemeen kan worden verwacht dat hoe meer het systeem afwijkt van de A-schematisatie, hoe meer het model gaat afwijken, bijvoorbeeld doordat een verandering in de afvoerverdeling tussen zomer- en winterbed na een rivierverruiming een hogere waarde geeft aan de kalibratieparameters van het zomerbed (de Jong 2021).

Het doel van een **Γ -kwadranttest** is om te testen of de voorspelonzekerheid gelijk blijft onder andere omstandigheden. Deze test is ook wel bekend als de *differential split-sample test* (Klemeš 1986b). Men voert hem uit door de splitsing niet willekeurig te doen, maar gespleten langs een bepaalde variabele. Men kan het model bijvoorbeeld kalibreren op lage afvoer, maar testen op hoge afvoer⁷.

Het doel van een **Δ -kwadranttest** is om te testen of de voorspelonzekerheid gelijk blijft onder andere omstandigheden in een veranderd onderliggend systeem. Het is hiermee een combinatie van de B- en Γ -kwadrantstesten.

⁷ Het nadeel van deze test is wel dat men hierdoor een mogelijkheid lijkt te verliezen om het model te verbeteren voor hoge afvoeren. Immers, men kan niet testen op een dataset waarop ook gekalibreerd is. Kalibreren op alle data zorgt voor hogere nauwkeurigheid tijdens kalibratie, maar hiermee verliest men het vermogen om te testen op nauwkeurigheid. Het ontwerpen van modellen die extrapoleren zonder significant verlies van nauwkeurigheid, waardoor kalibratie op alle data niet nodig is, verdient daarom aanbeveling.

3.2.3 Het berekenen van de modelonzekerheid

Modelonzekerheid is onzekerheid in modeluitvoer, door variatie (onzekerheid) in modelinstellingen (Figuur 3.1). Maar hoe kan deze variatie bepaald worden? Er zijn hiervoor, in grote lijnen, twee aanpakken.

3.2.3.1 Voorwaartse analyse

De conceptueel eenvoudigste methode is **voorwaartse analyse** (EN: *forward analysis*). Hierbij wordt aangenomen dat de variatie in modelinvoer bekend is (waarbij 'modelinvoer' breed kan worden opgevat, en in principe alles in Tabel 1.1 kan omvatten), en dat op basis daarvan de variatie in modeluitvoer berekend moet worden. De variatie in modelinvoer wordt bijvoorbeeld berekend uit expertanalyse of veldmetingen (Jord J Warmink et al. 2011). Voor riviermodellen is dit doorgaans een parameter die invloed heeft op de ruwheid, zoals de sediment- of plantdiameters. Om te berekenen hoe variatie in modeluitvoer doorwerkt op modeluitvoer, is **Monte Carlo analyse** de enige bekende universeel toepasbare methode⁸.

Het nadeel van voorwaartse analyse is dat er geen gebruik wordt gemaakt van metingen aan modeluitvoer, zoals waterstanden. Hierdoor wordt er geen selectie uitgevoerd op combinaties van mogelijke modelinvoeren die leiden tot een goede overeenkomst met metingen (in Bayesiaanse termen: de variatie wordt niet geüpdatet met de informatie uit metingen). Daardoor is deze methode vooral geschikt in situaties waar geen of geen betrouwbare metingen zijn, en men enkel op een model moet vertrouwen.

3.2.3.2 Inverse analyse

De andere aanpak is **inverse analyse** (EN: *inverse problem analysis*). Hierbij wordt aangenomen dat de modeluitvoer bekend is, maar de variatie in modelinvoer niet. Hier onderscheiden we twee abstractieniveaus.

Op het eerste niveau optimaliseren we modelinstellingen naar een enkele waarde, d.w.z. 'traditionele' of **deterministische kalibratie**. De modelonzekerheid is dan per definitie 'nul'. Op het tweede niveau worden modelinstellingen niet behandeld als een enkele waarde, maar als verdelingen. Die verdeling beschrijft de kans dat een bepaalde waarde leidt tot goede modeluitkomsten. We spreken dan over **probabilistische kalibratie**⁹.

In het algemeen kan men stellen dat hoe meer metingen⁶ beschikbaar zijn, hoe kleiner de modelonzekerheid is die door inverse analyse wordt bepaald. Modelonzekerheid kan dus verkleind worden door meer metingen te doen. Voorspelonzekerheid kan enkel verkleind worden door een beter model te maken of door betere metingen te doen.

Tot slot merken we bij het bovenstaande op dat modelonzekerheid en voorspelonzekerheid niet los van elkaar bepaald kunnen worden. Wanneer men deterministisch kalibreert, wordt de voorspelonzekerheid bepaald voor *dat* gekalibreerde model. Wanneer men probabilistisch kalibreert, is voorspelonzekerheid onderdeel van die berekening. In alle gevallen spreken we dus over voorspelonzekerheid, *gegeven* de modelonzekerheid.

⁸ D.w.z.: Monte Carlo (MC) analyse zal, wanneer men genoeg simulaties doet, altijd leiden tot het juiste antwoord. MC is echter wel erg rekenintensief. Er bestaan (benaderings)methodes die efficiënter zijn dan MC, maar die niet altijd leiden tot het juiste antwoord of slechts tot een (eerste) order benadering.

⁹ Methodes hiervoor zijn bijvoorbeeld *maximum likelihood estimation* of *Markov-Chain Monte Carlo*.

3.2.3.3 Praktische toepassing

In de praktijk is rekentijd een belangrijke beperking op de kalibratie of onzekerheidskwantificatiemethoden die voorhanden zijn. Inverse analyse kost enorm veel berekeningen en is praktisch onmogelijk wanneer een enkele berekening langer dan een paar seconden duurt. Tabel 3.1 geeft een overzicht van de orde grootte van het aantal berekeningen gebaseerd op de in dit rapport genoemde literatuur.

Om deze redenen wordt in de praktijk vaak volstaan met het afleiden van de voorspelonzekerheid in combinatie met deterministische kalibratie. In onderzoeksprojecten wordt om dezelfde reden voor toepassingen die diep in het bereik liggen waar modelonzekerheid waarschijnlijk domineert (Figuur 3.3), voorwaartse analyse gebruikt. Het gebruik van inverse technieken zoals MCMC vindt wel toepassing in datamodellen zoals QH-relaties.

Tabel 3.1 Rekenlast van de verschillende methodes

Methode	Orde grootte van # berekeningen in riviermodellen
Voorwaartse analyse	Honderden
Deterministische kalibratie	Tientallen
Probabilistische kalibratie	Honderden tot duizenden (onbekend)
Markov-Chain Monte Carlo	Duizenden tot tienduizenden

Een andere praktische voetnoot is dat modelonzekerheid veronderstelt dat sommige instellingen onzeker zijn. Welke instellingen dat zijn, moet door de modelleur bepaald worden. Een formele kwantitatieve methode om dat vast te stellen is **gevoeligheidsanalyse**. Een kwalitatieve methode is expert ondervraging (*EN: expert elicitation*), zie bijvoorbeeld (Jord J Warmink et al. 2011).

3.3 Een standaard modelbijlage voor riviermodellen

3.3.1 Riviermodellen van Rijkswaterstaat

Rijkswaterstaat heeft aan Deltares gevraagd om uit te werken waarom en hoe onzekerheid een rol speelt bij haar stromingsmodellen, en daarbij te komen tot een praktische basis om model- en voorspelonzekerheid in de praktijk te beoordelen.

De stromingsmodellen van Rijkswaterstaat zijn 1D en 2D modellen die worden gebouwd en onderhouden binnen het KPP (Kennis voor de Primaire Processen) raamwerk. De schematisatie van deze modellen wordt op een gestandaardiseerde manier afgeleid van een geodatabase (*Baseline*). Modellen worden met een bepaalde frequentie *geactualiseerd*, waarbij veranderingen in het fysieke systeem (zoals bodemligging, rivierbed, helling, begroeiing, aanwezig infrastructuur, etc.) worden overgenomen in de modellen. Daarom zijn er bijvoorbeeld modellen die de situatie in 1995 beschrijven, en modellen die de situatie in 2011 beschrijven.

De modellen worden deterministisch gekalibreerd op basis een subset van de beschikbare data, terwijl getest wordt met een andere subset (de Jong and Yossef 2021). Deze aanpak is het best te vergelijken met kruisvalidatie - in termen van de onzekerheidsmatrix (Figuur 3.3) is dit een A-kwadrant test (Figuur 3.4). Omdat schematisaties van verschillende jaren worden gebruikt, zijn er wel elementen van een stationariteitstest (B-kwadrant) aanwezig. Echter, de manier van toetsing en het splitsen van data zijn vaak niet expliciet gericht op een stationariteitstest. Daarbij moet wel worden opgemerkt dat het nog een openstaande vraag is hoe men de keuze voor het splitsen van data moet uitvoeren.

Het is bijvoorbeeld niet duidelijk hoeveel verandering van het riviersysteem leidt tot verandering van de modelnauwkeurigheid, hoe men 'verandering van het riviersysteem' überhaupt kan kwantificeren (zie aanbeveling 1d in paragraaf 5.2).

Omdat modellen niet expliciet getest worden voor toepassingen in het Γ -kwadrant (extrapolatie) of B-kwadrant (stationariteit), is de voorspelonzekerheid voor sommige toepassingen onbekend:

- Wat is de verwachte waterstand bij een extreme afvoer?
- Welke garantie kan worden afgegeven dat de verwachte vaardiepte voor scheepvaart minimaal 2.8m is bij minimale afvoer?
- Hoe groot is het effect van een ingreep op de waterstand?

Er zijn desalniettemin wel studies gedaan met Rijkswaterstaatmodellen die dergelijke vragen verkend hebben. Zo is de modelonzekerheid bij de maatgevende afvoer op basis van voorwaartse analyse voor de Waal berekend in (Koen D Berends et al. 2018) voor een D-Hydro model en (J. Warmink 2011) voor een WAQUA model.

3.3.2 Doelstelling van een modelbijlage

De doelstelling van een modelbijlage is het inzichtelijk ontsluiten van beschikbare informatie over modelnauwkeurigheid- en onzekerheid voor gebruikers, door het opstellen van een gestandaardiseerd formulier dat inhoudelijk verbonden is aan het terminologisch raamwerk van hoofdstuk 2.

3.3.3 Onderdelen van de modelbijlage

Om tot een geordende opsomming van informatie te komen, stellen we de volgende onderdelen voor

- Een modelinformatieformulier, waarbij ten minste wordt opgenomen:
 - De naam en versie van het model
 - De software waarmee het model is getest
 - Referenties naar de documentatie van het model
- Een overzicht van testen, waarbij per ten minste wordt opgenomen:
 - Welke uitvoervariabele getest/gekwantificeerd word
 - Voor welk gebruik getest wordt, gelinkt aan de kwadranten van Figuur 3.3.
 - Op welk bereik getoetst is en met welke dataset
 - De uitkomst van de test

4 Toepassing op een casus

4.1 Het Maas-model

4.1.1 Casus & documentatie

De casus is het zesde-generatie 2D model van de Maas. Dit model is opgezet in de D-Hydro software. De bouw, kalibratie en validatie van het model is beschreven in (de Jong and Yossef 2021). Aanvullende testen zijn gedaan door (de Jong 2021). Tot slot is een actualisatie uitgevoerd (van der Deijl, de Jong, and Visser 2021). Op deze drie rapporten baseren wij de beschrijving van de casus in dit rapport.

4.1.2 Modelinformatieformulier

De definitie van het model is vastgesteld in paragraaf 1.1 als “een familie van versie-, gebieds- en softwarespecifieke modelschematisaties”. Binnen deze definitie vallen verschillende *schematisaties* die elk een verschillende staat van het riviersysteem beschrijven. Voor het Maasmodel zijn dit onder meer:

- dflowfm2d-maas-j93_6-v1a;
- dflowfm2d-maas-j95_6-v1a;
- dflowfm2d-maas-j10_6-v1a;
- dflowfm2d-maas-j14_6-v1a;
- dflowfm2d-maas-j19_6-v2a;
- dflowfm2d-maas-beno19_6-v1a.

Al deze *schematisaties* (die in de praktijk ook vaak als “model” worden aangeduid) worden hier dus gezien als instanties van het “hoofdmodel” voor een specifiek configuratie. Deze schematisaties zijn allemaal onderdeel van dezelfde generatie modellen, oftewel dezelfde modelfamilie. Uitgangspunten en kalibraties worden binnen een generatie zoveel mogelijk gelijk gehouden, waardoor het testoverzicht ook voor een volledige generatie kan gelden. De invloed van significante veranderingen aan modellen of methodiek, die leidt tot een verhoging van het versienummer binnen een generatie, op de overdraagbaarheid van de testresultaten moet dan worden aangetoond – bijvoorbeeld door verschiltesten uit te voeren.

Voor de naamgevingsconventie verwijzen wij naar de memo “Naamgeving conventies modellen Rijkswaterstaat versie 2.0”. Een overzicht van de schematisaties wordt bijgehouden in de Factsheet Maas G6.

Op basis van bovenstaande vullen wij het modelinformatieformulier als volgt in:

Tabel 4.1 Modelinformatieformulier van dflowfm2d-maas_6. Deze informatie is ook (deels) opgenomen in de zgn. 'factsheets'.

Algemene informatie	
Naam	dflowfm2d-maas_6*
Versie	1a
Te gebruiken met software versie(s)	D-HYDRO 2021-04 of hoger***
Gebaseerd op database	De baseline-maas_6* familie
Randvoorwaardemodellen	RGWM 2.2.1 Qf relatie Eijsden**
Documentatie	Ontwikkeling: Deltares rapport 11200569-003-ZWS-0014 Actualisatie j19: 11206813-002-ZWS-0021

* In de naamgevingsconventie is (nog) geen regel opgenomen voor de deze onderdelen. Het wordt aangeraden om hier een conventie voor op te nemen.

** Voor deze relatie is geen versienummering bekend

*** De testen kunnen per deelmodel met andere softwareversie uitgevoerd zijn. Hier hebben we de laatste versie opgenomen waarmee dat gebeurd is. Het verdient aanbeveling om hier een versie op te nemen waarmee vastgesteld kan worden dat alle testen herhaald kunnen worden met het zelfde resultaat. Indien dat niet mogelijk is, bijvoorbeeld door grote softwarematige veranderingen, verdient dat speciale aandacht.

4.1.3 Testoverzicht

Uit de drie rapporten zijn in twee rapporten (van der Deijl, de Jong, and Visser 2021; de Jong and Yossef 2021) testen uitgevoerd die uitspraken doen over de modelnauwkeurigheid (voorspelonzekerheid). Deze testen zijn geïnterpreteerd in het kader van de onzekerheidsmatrix (Figuur 3.3) en samengevat in Tabel 4.2. De details van de testen zijn opgenomen in bijlage A. Tabel 4.2. is bedoeld als een snelle referentie: als een gebruiker geïnteresseerd is in afwijkende variabelen (zoals stroomsnelheden), ziet men meteen dat hier niet op getest is. Als de gebruiker geïnteresseerd in waterstanden, kan hij met deze tabel vinden welke hij nader moet bekijken.

Tabel 4.2 Overzicht van de testen met het Maasmodel. Het verschil tussen een test en een verkenning wordt in de onderstaande paragraaf benoemt. De details van de testen zijn opgenomen in bijlage A.

Test id	Kwadrant	Omschrijving en testvariabele	Type onzekerheid
#1	A-kwadrant test	Waterstand op LMW stations bij lage afvoer	Voorspelonzekerheid
#2	A-kwadrant test	Waterstand op LMW stations bij middellage afvoer	Voorspelonzekerheid
#3	B-kwadrant test	Waterstand op LMW stations bij middelhoge afvoer	Voorspelonzekerheid
#4	B-kwadrant test	Waterstand op LMW stations bij hoge afvoer	Voorspelonzekerheid
#5	A-kwadrant test	Waterstand op LMW stations bij zeer hoge afvoer	Voorspelonzekerheid
#6	Δ -kwadrant verkenning	Waterstand op LMW stations bij extreme afvoer	Voorspelonzekerheid
#7	Γ -kwadrant verkenning	Waterstand op LMW stations bij extreme afvoer	Voorspelonzekerheid

Uit het opstellen van deze lijst kwamen de volgende aandachtspunten naar voren:

- Er moet een onderscheid gemaakt worden tussen een test (waarbij de uitspraken gebaseerd zijn op vergelijking met observaties) en een verkenning (waarbij de uitspraken gebaseerd zijn op bepaalde aannames, maar niet getoetst met metingen). We zien dan dat er vooral in het A- en B-kwadrant testen zijn uitgevoerd. De andere twee kwadranten zijn wel verkend, maar niet getest. Mogelijk moet deze classificering nog worden uitgebreid met bijvoorbeeld expertbeoordelingen, hoewel dit voor deze casus op dit moment niet aan de orde is.
- Tijdens de modelbouw worden alle mogelijke afvoerbereiken gebruikt bij modelkalibratie. Een goede extrapolatie (Γ -kwadrant) test hebben is daardoor niet beschikbaar.
- Er zijn testen die in de kalibratierapportage die geclassificeerd kunnen worden als een B-kwadrant (stationariteits)test (de Jong and Yossef 2021). Dit wordt echter niet expliciet zo benoemd, en was ook niet het doel: *“bij de keuze voor kalibratieperioden wordt gekeken naar de stabiliteit van de afvoer, de aansluiting van metingen en schematisatie, de mogelijke negatieve invloed van niet-gemodelleerde of onvolkomen weergegeven processen, en een bij voorkeur zo recent mogelijke periode”*. Omdat er niet expliciet op stationariteit wordt getest, is er ook geen afweging geweest over de geschiktheid van de keuzes voor testperioden. Zo'n afweging zou bijvoorbeeld kunnen bestaan uit het beschouwen van de veranderingen die het systeem heeft ondergaan, en die zouden kunnen leiden tot nauwkeurighedsverlies. Een aanzet om dit wel expliciet mee te nemen wordt gegeven in aanbeveling 5 van (de Jong 2021), die in dit rapport is overgenomen.
- Sommige testen van verschillende schematisaties (deelmodellen) zijn uitgevoerd met verschillende softwareversies. Het verdient vanuit het oogpunt van duidelijkheid aanbeveling om te onderzoeken of er per familie één advies kan worden gegeven voor de te gebruiken software.

5 Discussie, conclusies en aanbevelingen

5.1 Discussie

5.1.1 Het gebruik van modellen voor verschilanalyse

In dit rapport hebben we ons enkel beperkt tot de onzekerheid van modeluitvoer (zie paragrafen 0 en 3.1) van een enkele modelsimulatie. In de praktijk wordt deze uitvoer echter ook vaak gebruikt in ingreepeffectstudies of verschilanalyses.

Vershilanalyses bestaan uit het bepalen van het verschil tussen twee modelsimulaties. Vaak wordt dit toegepast wanneer men wil weten wat het effect is van een maatregel of een andere verandering in het systeem. Daarbij wordt doorgaans aangenomen dat de onzekerheid van verschilanalyses aanzienlijk kleiner is dan de onzekerheid die kan bestaan in de 'absolute' modeluitvoer (Mosselman 2018). Dit is waarschijnlijk een juiste aanname, omdat de modeluitvoer van twee gerelateerde modellen gecorreleerd is (Koen D Berends et al. 2018). Hoe onzeker een verschilanalyse is, is niet eenvoudig te berekenen, omdat deze correlatie in de praktijk onbekend is, en omdat empirische metingen om de voorspelonzekerheid van verschilanalyses in de praktijk te toetsen ontbreken (Koen D. Berends et al. 2021).

5.1.2 Bronnen van onzekerheid en relatie tot andere raamwerken

In dit rapport wordt onzekerheid beschreven als een bandbreedte van of rondom modeluitvoer (Figuur 3.1). We hebben uit praktisch oogpunt gekozen voor deze in sommige opzichten nauwe definitie, omdat dit goed aansluit bij de huidige praktijk. Hier willen we kort toelichten hoe deze keuze zich verhoudt tot andere raamwerken voor onzekerheid en validatie, specifiek die van (Walker et al. 2003) en (Rykiel 1996).

5.1.2.1 Verhouding met onzekerheidsbronnen en onzekerheidsraamwerken

(Walker et al. 2003) stelde een classificatie van onzekerheid voor die zich specifiek richtte op modellen. Zij onderscheiden vijf verschillende locaties van onzekerheid: *Context*, *Model uncertainty*, *Inputs*, *Parameters* en *Model outcome uncertainty*. De onzekerheid die in dit rapport als 'Totale onzekerheid' wordt benoemt is in termen van Walker et al. *model outcome uncertainty*, die zij definiëren als "De accumulatie van onzekerheid in de modeluitvoer die voor beslissingen relevant zijn". Die accumulatie komt voort uit de andere locaties, die wij in dit rapport niet expliciet benoemen, maar wel belangrijk kunnen zijn in de interpretatie van de onzekerheid. Bijvoorbeeld als men wil weten *waarom* de onzekerheid toeneemt bij extrapolatie: komt dit door onzekerheid in de afvoer (*Inputs*) of door een onvolkomenheid in het model (*model uncertainty*)? Deze uitsplitsing hebben we in dit rapport niet gemaakt.

Een ander onderscheid dat Walker et al. (2003) maken is in *niveaus* van onzekerheid. Hier onderscheiden zij statistische onzekerheid, scenario onzekerheid, erkende onwetendheid en totale onwetendheid. Dit rapport focust zich nadrukkelijk op statische onzekerheid, wat door de auteurs wordt omschreven als "alle onzekerheid die zich voldoende laat beschrijven in statische termen".

Een laatste onderscheid dat Walker et al. (2003) maken is in de aard van de onzekerheid, waar ze twee smaken erkennen: onzekerheid door gebrek van kennis (*epistemic uncertainty*) en inherente onzekerheid (*variability uncertainty*). Wij maken geen onderscheid hierin: alle kwantificeerbare onzekerheid wordt bijeengenomen, ongeacht de aard van de bron.

Er zijn meer raamwerken dan die van Walker et al. (2003), zie bijvoorbeeld (Skinner, Rocks, and Pollard 2014), die meestal specifiek zijn gericht op het type model en het jargon van de modelgebruikers. Consensus over een raamwerk dat alles vangt is er niet, en hoe over onzekerheid wordt gecommuniceerd (de *framing*) is in de praktijk misschien belangrijker (Guillaume et al. 2017).

5.1.2.2 Modelvalidatie, datavalidatie en verificatie

De termen 'validatie' en 'verificatie' worden vaak door elkaar gebruikt (Oreskes 2004). In dit rapport gebruiken we 'validatie' als de activiteit om voorspelonzekerheid vast te stellen. In deze paragraaf willen we aan de hand van het raamwerk van Rykiel (1996) stil staan bij andere types validatie.

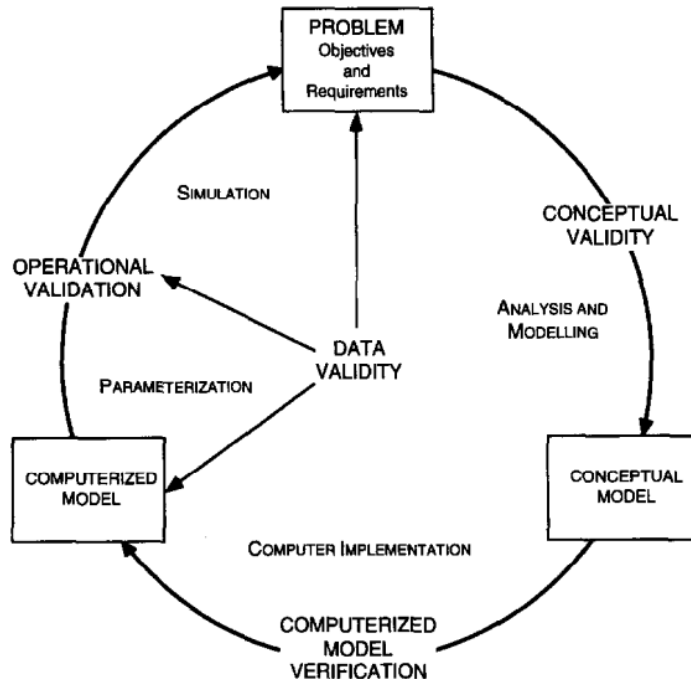
Rykiel (1996) maakte een onderscheid tussen verschillende manieren van valideren: conceptueel valideren, datavalidatie, software verificatie en operationeel valideren (Figuur 5.1). De testen die in dit rapport worden genoemd vallen in deze context onder *operationele validatie*.

De testen die hieraan voorafgaan zijn desalniettemin wel van belang. Onder datavalidatie valt bijvoorbeeld het controleren op uitschieters, gaten, en apparatuurfalen. Dit is dus belangrijk voor het afleiden van randvoorwaarden en de gebiedsbeschrijving. Conceptuele validatie betreft het toetsen van basisprincipes aan de werkelijkheid, zoals een nieuwe overlaatformule voor een kribben, of een nieuwe vegetatieweerstandsfomule. Verificatie is de test of dat conceptuele idee goed is geïmplementeerd in de software.

Operationele validatie zijn afhankelijk van de testen die eraan vooraf gaan. Als het concept van een overlaat niet klopt, heeft dat invloed op de operationele validiteit. Idem, als er een bug zit in de implementatie. In statische termen kunnen we zeggen dat de onzekerheden die uit de testen in dit rapport volgen *conditioneel zijn op* (afhankelijk zijn van) alle aannames die hieraan voorafgaan. Het is om deze reden belangrijk om deze condities expliciet te vermelden. In de casus dit in dit rapport besproken zijn, zijn dit bijvoorbeeld:

- De methode die gebruikt wordt om de randvoorwaarden te bepalen (bijvoorbeeld $Q(f)$ relaties, zie Figuur 1.1).
- De onderliggende gebiedsdatabase en de methode om deze te onderhouden.
- Het versienummer van de gebruikte modelsoftware.

Een significante verandering aan deze voorliggende methodes kan ertoe leiden dat de testresultaten niet meer relevant zijn. Enkele uitzonderingen daargelaten, is de aanlevering van de onderliggende methodes in de besproken casus vastgelegd en voorzien van versiebeheer, wat controle op die afhankelijkheid waarborgt.



Figuur 5.1 De modelcyclus volgens Rykiel (1996).

5.1.3 Opname van derden studies

Tijdens de workshop van 30 november 2021 kwam de vraag op of studies van derden opgenomen moeten worden in het testoverzicht. Een voorbeeld hiervan is de studie van (J J Warmink, Straatsma, et al. 2013), die modelonzekerheid bij extreme afvoer op de Waal kwantificeren met behulp van een WAQUA model. Uit deze discussie volgde geen fundamenteel bezwaar, maar wel een voorbehoud op review door Rijkswaterstaat. Het standaardiseren van de validatieprocedure en identificeren van witte vlekken (erkende onwetendheid) kan bijdragen aan valorisatie van derden studies om te komen tot een beter begrip van de onzekerheid van modellen.

5.2 Conclusie en aanbevelingen

Rijkswaterstaat heeft aan Deltares gevraagd om uit te werken waarom en hoe onzekerheid een rol speelt bij haar stromingsmodellen, en daarbij te komen tot een praktische basis om onzekerheid in de praktijk te beoordelen.

Hieraan is invulling gegeven door, op basis van literatuur, een conceptuele basis voor te stellen die specifiek is gericht op de gebruikspraktijk van Rijkswaterstaat-stromingsmodellen. Deze conceptuele basis is vervolgens geconcretiseerd door een modelbijlage te ontwerpen en die toe te passen op de casus “zesde-generatie 2D model voor de Maas”.

De doelstelling van een modelbijlage is het inzichtelijk ontsluiten van beschikbare informatie over modelnauwkeurigheid- en onzekerheid voor gebruikers, door het opstellen van een gestandaardiseerd formulier dat inhoudelijk verbonden is aan het terminologisch raamwerk van hoofdstuk 2.

Uit deze casus bleek dat in de praktijk alleen gekeken wordt naar voorspelonzekerheid (modelnauwkeurigheid), die wordt uitgedrukt in statische momenten (bias en standaardafwijking). De onzekerheid van alle drie toepassingsgebieden (gamma, beta en delta kwadrant) zijn benoemd in de beschouwde rapportages.

Van de vier testen wordt enkel de kruisvalidatietest (A-kwadrant) en stationariteitstest (B-kwadrant) kwantitatief getoetst aan de hand van metingen. De extrapolatie (Γ -kwadrant) en stationariteit+extrapolatie (Δ -kwadrant) zijn wel verkend, maar niet getoetst.

De Maas-casus heeft hiermee geresulteerd in een model waarvan in elk van de kwadranten kennis beschikbaar is over de verwachte nauwkeurigheid van het model. Uit deze casus blijkt dat expliciete benoeming van het doel van validatie ontbreekt en hiermee ook een beschouwing van de voor die toepassing benodigde nauwkeurigheid. Dit bemoeilijkt de interpretatie van de waarde van een test. Uit deze casus volgen de volgende aanbevelingen:

1. Standaardiseer de beschrijving van validatietesten in rapportages.
 - a Benoem in kalibratie- en validatierapporten expliciet het doel en eventueel criteria van de test, bijvoorbeeld het testen van het vermogen van het model om te extrapoleren naar hogere afvoeren, of het vermogen om effecten van gebiedsveranderingen te kwantificeren.
 - b Beschrijf hoe de gekozen methode aansluit bij het doel.
 - c Overweeg om de te gebruiken software niet per deelmodel, maar per modelfamilie vast te stellen, uit het oogpunt van duidelijkheid.
 - d Voor stationariteitstesten is het aan te raden om te onderzoeken of er een onafhankelijke maat kan worden gevonden voor de mate van gebiedsverandering. Een aanzet hiervoor is gegeven in (de Jong 2021) en (K D Berends, Warmink, and Hulscher 2018). Zo'n beschouwing kan ook richting geven aan de vraag hoeveel gebiedsverandering genoeg is om te kunnen spreken over een stationariteitstest.
2. Voer extra testen uit om de 'witte vlekken' af te dekken
 - a Een goede extrapolatie (Γ -kwadrant) test hebben is niet beschikbaar, omdat op alle verschillende meetbare afvoeren wordt gekalibreerd. Het verdient aanbeveling om te onderzoeken hoe er toch een kwantitatieve maat kan worden gegeven voor het vermogen van een model om te extrapoleren naar hogere afvoeren.
 - b Het hoogwater van juli 2021 op de Maas geeft een mogelijkheid om een gecombineerde extrapolatie-stationariteitstest (Δ -kwadrant) uit te voeren, omdat de afvoer (bovenstrooms) hoger is dan waar het model op is gekalibreerd, en het jaar (en dus de schematisatie) recenter dan waar het model op is gekalibreerd.
3. In de naamgevingsconventie is er nog geen conventie opgenomen voor het benoemen van de 'modellenfamilie', waarvan schematisaties onderdeel zijn. Het verdient aanbeveling om dit op te nemen.
4. De generatie van de bovenstroomse randvoorwaarden met behulp van het Q(f)-model is een van de weinige onderdelen waarvan geen versienummering wordt bijgehouden of gedocumenteerd. Het verdient aanbeveling om ook dit model, waar de relevantie van de testen kritisch van afhankelijk zijn, op te nemen in een gestandaardiseerd versiebeheersysteem.
5. Evalueer deze casus en discussieer over eventuele gezamenlijke uitbouw van de conceptuele basis.
 - a Deze modelbijlage vormt een eerste aanzet naar een gestandaardiseerde manier om te onzekerheid in de praktijk te beoordelen in het kader van stromingsmodellen van Rijkswaterstaat. Het komen tot een gedragen, gezamenlijk begrip is belangrijk. Het is daarom aan te raden deze casus te evalueren met relevante belanghebbenden.

- b. Er is enige overlap met de 'factsheets', die worden gemaakt bij oplevering van schematisaties. Het verdient aanbeveling om op termijn de beschikbare informatie over een model te bundelen. Mogelijk kan de in dit rapport aanbevolen tabellen worden gezien als een uitbreiding van de reeds bestaande factsheets.
- c. Deze basis is met opzet nauw gekozen om aan te sluiten bij de huidige praktijk, waardoor sommige vragen (zoals "wat is de onzekerheid van verschilanalyses") niet binnen dit kader beantwoord kunnen worden. Het verdient aanbeveling om te evalueren hoe deze vragen een plaats kunnen krijgen in een toekomstige, bredere conceptuele basis.

Referenties

- Asselt, M B A van, R Langendonk, F van Asten, A van der Giessen, P H M Janssen, P S C Heuberger, and I Geuskens. 2001. "Uncertainty & RIVM's Environmental Outlooks; Documenting a Learning Process." <https://www.rivm.nl/bibliotheek/rapporten/550002001.pdf>.
- Becker, Anke. 2012. "Rijn-Modellen 5de Generatie: Modelmodel, Kalibratie En Verificatie WAQUA."
- Berends, K. 2019. "Human Intervention in Rivers." Enschede, The Netherlands: University of Twente. <https://doi.org/10.3990/1.9789036548823>.
- Berends, K D, J J Warmink, and S.J.M.H. Hulscher. 2018. "Efficient Uncertainty Quantification for Impact Analysis of Human Interventions in Rivers." *Environmental Modelling {&} Software* 107 (September): 50–58. <https://doi.org/10.1016/j.envsoft.2018.05.021>.
- Berends, Koen D., Matthijs R. A. Gensen, Jord J. Warmink, and Suzanne J. M. H. Hulscher. 2021. "Multidecadal Analysis of an Engineered River System Reveals Challenges for Model-Based Design of Human Interventions." *CivilEng* 2 (3): 580–98. <https://doi.org/10.3390/civileng2030032>.
- Berends, Koen D., Un Ji, W.E. (Ellis) Penning, Jord J. Warmink, Joongu Kang, and Suzanne J.M.H. Hulscher. 2020. "Stream-Scale Flow Experiment Reveals Large Influence of Understory Growth on Vegetation Roughness." *Advances in Water Resources* 143 (September): 103675. <https://doi.org/10.1016/j.advwatres.2020.103675>.
- Berends, Koen D, Menno W Straatsma, Jord J Warmink, and Suzanne J M H Hulscher. 2018. "Uncertainty Quantification of Flood Mitigation Predictions and Implications for Decision Making." *Natural Hazards and Earth System Sciences Discussions*, November, 1–25. <https://doi.org/10.5194/nhess-2018-325>.
- Beven, Keith. 2016. "Facets of Uncertainty: Epistemic Uncertainty, Non-Stationarity, Likelihood, Hypothesis Testing, and Communication." *Hydrological Sciences Journal* 61 (9): 1652–65. <https://doi.org/10.1080/02626667.2015.1031761>.
- Breedveld, Willem. 1999. "Dank de Klokkenluider." *Trouw*, 1999. <https://www.trouw.nl/nieuws/dank-de-klokkenluider~b49d5db9/>.
- Deijl, Eveline van der, Jurjen de Jong, and Ton Visser. 2021. "Actualisatie Zesde-Generatie Maas-Modellen. Deltares Rapport 11206813-002-ZWS-0021."
- Domhof, Boyan C A, Koen D Berends, Aukje Spruyt, Jord J Warmink, and Suzanne J M H Hulscher. 2018. "Discharge and Location Dependency of Calibrated Main Channel Roughness: Case Study on the River Waal." Edited by A Paquier and N Rivière. *E3S Web of Conferences* 40: 6038. <https://doi.org/10.1051/e3sconf/20184006038>.
- Guillaume, Joseph H A, Casey Helgeson, Sondoss Elsayah, Anthony J Jakeman, and Matti Kummu. 2017. "Toward Best Practice Framing of Uncertainty in Scientific Publications: A Review of Water Resources Research Abstracts." *Water Resources Research* 53 (8): 6744–62. <https://doi.org/10.1002/2017wr020609>.
- Jong, Jurjen de. 2021. "Advisering in de Keuze van Kalibratiefactoren Voor MHW-Afvoeren. Deltares Rapport 11206813-002-ZWS-0012."

- Jong, Jurjen de, and Mohamed Yossef. 2021. "Ontwikkeling Zesde-Generatie Maas-Model. Deltares Rapport 11200569-003-ZWS-0014-v1.1."
- Klemeš, V. 1986a. "Dilettantism in Hydrology: Transition or Destiny?" *Water Resources Research* 22 (9): 177S-188S. <https://doi.org/10.1029/WR022i09Sp0177S>.
- . 1986b. "Operational Testing of Hydrological Simulation Models." *Hydrological Sciences Journal* 31 (1): 13–24. <https://doi.org/10.1080/02626668609491024>.
- Mosselman, Erik. 2018. "Modelling in Applied Hydraulics: More Accurate in Decision-Making Than in Science?" In *Advances in Hydroinformatics*, 741–49. Springer Singapore. https://doi.org/10.1007/978-981-10-7218-5_53.
- Mosselman, Erik, and Thai Binh Le. 2016. "Five Common Mistakes in Fluvial Morphodynamic Modeling." *Advances in Water Resources* 93 (July): 15–20. <https://doi.org/10.1016/j.advwatres.2015.07.025>.
- Oreskes, Naomi. 2004. "Science and Public Policy: What's Proof Got to Do with It?" *Environmental Science & Policy* 7 (5): 369–83. <https://doi.org/10.1016/j.envsci.2004.06.002>.
- Petersen, A C, P H M Janssen, J P van der Sluijs, J S Risbey, and J R Ravetz. 2003. "RIVM/MNP Guidance for Uncertainty Assessment and Communication." Bilthoven. https://www.rivm.nl/bibliotheek/digitaaldepot/Guidance_MC_QS-Q.pdf.
- Rykiel, Edward J. 1996. "Testing Ecological Models: The Meaning of Validation." *Ecological Modelling* 90: 229–44. [https://doi.org/10.1016/0304-3800\(95\)00152-2](https://doi.org/10.1016/0304-3800(95)00152-2).
- Saltelli, Andrea, Gabriele Bammer, Isabelle Bruno, Erica Charters, Monica Di Fiore, Emmanuel Didier, Wendy Nelson Espeland, et al. 2020. "Five Ways to Ensure That Models Serve Society: A Manifesto." *Nature* 582 (7813): 482–84. <https://doi.org/10.1038/d41586-020-01812-9>.
- Skinner, Daniel J C, Sophie A Rocks, and Simon J T Pollard. 2014. "A Review of Uncertainty in Environmental Risk: Characterising Potential Nature, Locations and Levels." *Journal of Risk Research* 14 (2): 195–219. <https://doi.org/10.1080/13669877.2013.794150>.
- Stein, Seth, and Jerome L. Stein. 2013. "Shallow Versus Deep Uncertainties in Natural Hazard Assessments." *Eos, Transactions American Geophysical Union* 94 (14): 133–34. <https://doi.org/10.1002/2013EO140001>.
- Straatsma, Menno W, Marcel van der Perk, Aafke M Schipper, Reinier J W de Nooij, Rob S E W Leuven, Fredrik Huthoff, and Hans Middelkoop. 2013. "Uncertainty in Hydromorphological and Ecological Modelling of Lowland River Floodplains Resulting from Land Cover Classification Errors." *Environmental Modelling & Software* 42: 17–29. <https://doi.org/10.1016/j.envsoft.2012.11.014>.
- Thirel, Guillaume, Vazken Andréassian, and Charles Perrin. 2015. "On the Need to Test Hydrological Models under Changing Conditions." *Hydrological Sciences Journal* 60 (7–8): 1165–73. <https://doi.org/10.1080/02626667.2015.1050027>.
- Trouw. 1999. "Milieu-Instituut Liegt En Bedriegt." *Trouw*, 1999. <https://www.trouw.nl/nieuws/milieu-instituut-liegt-en-bedriegt-bfb64b86/>.
- Twijnstra, J.J., M.R.A. Gensen, J.J. Warmink, S.J.M.H. Hulscher, F. Huthoff, and G. Horn. 2020. "Water Balance in the Dutch River Rhine and Rating Curve Uncertainty." In *River Flow 2020*, 2171–78. CRC Press. <https://doi.org/10.1201/b22619-304>.

- Walker, W E, P Harremoës, J Rotmans, J P van der Sluijs, M B A van Asselt, P Janssen, and M P Krayenbühl von Krauss. 2003. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." *Integrated Assessment* 4 (1): 5–17. <https://doi.org/10.1076/iaij.4.1.5.16466>.
- Warmink, J J. 2014. "Dune Dynamic and Roughness under Gradually Varying Flood Waves, Comparing Flume and Field Observations." *Advances in Geosciences* 39: 115–21. <https://doi.org/10.5194/adgeo-39-115-2014>.
- Warmink, J J, M J Booij, H van der Klis, and S.J.M.H. Hulscher. 2013. "Quantification of Uncertainty in Design Water Levels Due to Uncertain Bed Form Roughness in the Dutch Rivier Waal." *Hydrological Processes* 27: 1646–63. <https://doi.org/10.1002/hyp.9319>.
- Warmink, J J, M W Straatsma, F Huhoff, M J Booij, and S J M H Hulscher. 2013. "Uncertainty of Design Water Levels Due to Combined Bed Form and Vegetation Roughness in the Dutch River Waal." *Journal of Flood Risk Management* 6: 302–18. <https://doi.org/10.1111/jfr3.12014>.
- Warmink, Jord. 2011. "Unraveling Uncertainties: The Effect of Hydraulic Roughness on Design Water Levels in River Models." University of Twente.
- Warmink, Jord J, Hanneke van der Klis, Martijn J Booij, and Suzanne J M H Hulscher. 2011. "Identification and Quantification of Uncertainties in a Hydrodynamic River Model Using Expert Opinion Elicitation." *Water Resources Management* 25 (2): 601–22. <https://doi.org/10.1007/s11269-010-9716-7>.

A Testen

Test #1	
Doel	De nauwkeurigheid vaststellen waterstanden op LMW locaties
Type test	A-kwadrant test
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations
Testperiode	Juli 2014
Bereik	0 m ³ s ⁻¹ t/m 100 m ³ s ⁻¹
Testresultaat	Bias: 4.8 cm Standaarddeviatie: 9.1 cm
Documentatie	"Validatie L" 11200569-003-ZWS-0014, hoofdstuk 5

Test #2	
Doel	De nauwkeurigheid vaststellen waterstanden op LMW locaties
Type test	A-kwadrant test
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations
Testperiode	Januari 2015
Bereik	100 m ³ s ⁻¹ t/m 800 m ³ s ⁻¹
Testresultaat	Bias: 1.1 cm Standaarddeviatie: 4.0 cm
Documentatie	Test "M1" 11200569-003-ZWS-0014, hoofdstuk 5

Test #3	
Doel	De nauwkeurigheid vaststellen waterstanden op LMW locaties
Type test	B-kwadrant test
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations
Testperiode	November 2010 → December 1994 (<i>Kalibratieperiode → Validatieperiode</i>)
Bereik	800 m ³ s ⁻¹ t/m 1700 m ³ s ⁻¹
Testresultaat	Bias: 16.2 cm Standaarddeviatie: 15.2 cm
Documentatie	Test "M2" 11200569-003-ZWS-0014, hoofdstuk 5

Test #4	
Doel	De nauwkeurigheid vaststellen waterstanden op LMW locaties
Type test	B-kwadrant test
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations

Test #4	
Testperiode	Januari 2011 → Januari 1995 (Kalibratieperiode → Validatieperiode)
Bereik	1700 m ³ s ⁻¹ t/m 2250 m ³ s ⁻¹
Testresultaat	Bias: 12.9 cm Standaarddeviatie: 11.3 cm
Documentatie	Test "H1" 11200569-003-ZWS-0014, hoofdstuk 5

Test #5	
Doel	De nauwkeurigheid vaststellen waterstanden op LMW locaties
Type test	A-kwadrant test*
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations
Testperiode	December 1993
Bereik	800 m ³ s ⁻¹ t/m 1700 m ³ s ⁻¹
Testresultaat	Bias: 18.9 cm Standaarddeviatie: 18.8 cm
Documentatie	Test "H2" 11200569-003-ZWS-0014, hoofdstuk 5

*We hebben nog geen goed idee hoe ver twee periodes uit elkaar moeten liggen om een goede B-test uit te voeren. Omdat het verschil hier klein is (1 jaar en 2 maanden), noemen we het een A-test

Test #6	
Doel	Inschatting van de verwachte modelfout bij extreme afvoer
Type test	Δ-kwadrant verkenning*
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations
Testperiode	Januari 1995 → Beno situatie (Kalibratieperiode → Verkenningsperiode)
Bereik	2750 m ³ s ⁻¹ → 4000 m ³ s ⁻¹ (Kalibratiebereik → Testbereik)
Testresultaat	Beredeneerde modelfout: -10 cm en + 40 cm.
Documentatie	11206813-002-ZWS-0012, hoofdstuk 2.3

*Omdat niet wordt vergeleken met metingen, noemen we het een verkenning

Test #7	
Doel	Inschatting van de verwachte modelfout bij extreme afvoer
Type test	Γ-kwadrant verkenning*
Type onzekerheid	Voorspelonzekerheid van deterministisch model
Grootheid	Waterstanden op LMW stations
Testperiode	Januari 1995
Bereik	2750 m ³ s ⁻¹ → 4000 m ³ s ⁻¹ (Kalibratiebereik → Testbereik)
Testresultaat	Beredeneerde verschil: -0.3 à -0.05 m
Documentatie	11206813-002-ZWS-0012, hoofdstuk 2.1

*Omdat niet wordt vergeleken met metingen, noemen we het een verkenning

Deltares is een onafhankelijk kennisinstituut voor toegepast onderzoek op het gebied van water en ondergrond. Wereldwijd werken we aan slimme oplossingen voor mens, milieu en maatschappij.

Deltares

www.deltares.nl