

HMK
- handbok i mät- och kartfrågor

Geodatakvalitet

2017



Förord 2017

Detta är den tredje versionen av HMK-Geodatakvalitet. Efter 2017 års uppdatering består introduktionen till HMK av följande tre samverkande dokument:

- [HMK – Introduktion 2017](#)
- *HMK – Geodatakvalitet 2017* (detta dokument) som är ett referensverk beträffande kvalitetsfrågor.
- *Terminologi, principer och trender inom geodatakvalitet* – en kompletterande rapport inom HMK:s serie ”Tekniska rapporter” med kortnamnet [HMK-TR 2015:1](#).

I förhållande till den förra versionen (HMK-Geodatakvalitet 2015) så har version 2017 endast genomgått marginella förändringar. Framför allt har dokumentnamn, dokumentens årtal samt länkar uppdaterats och aktualiteten i refererade standarder har kontrollerats. Dessutom har termerna *georeferering* och *lägesosäkerhet* tydliggjorts, se slutet av kapitel 1.

Arbetet med dokumentet har – över åren – huvudsakligen utförts av Clas-Göran Persson, Lantmäteriet. Medverkat har även Thomas Lithén, Gunhild Lönnberg och Torsten Svärd, Lantmäteriet. En expertgranskning av HMK – Geodatakvalitet gjordes 2015.

Gävle 2017-09-30

Anders Grönlund
Uppdragsledare HMK

[Samlade Förord](#)

Innehållsförteckning

1	Inledning	5
2	Datakvalitet	9
2.1	Definition av geodatakvalitet	9
2.2	Datakvalitetsmodell och informationsmodellering	9
2.3	Datakvalitetsfilosofi	10
2.4	Kvalitetsmärkning och metadata	12
2.5	Dataproduktspecifikationen	13
2.6	HMK-standardnivåer	13
2.7	Kvalitetsteman och kvalitetsparametrar	15
2.8	Datakvalitet för bild- och laserdata	18
3	Kontroll av geodata	21
3.1	Olika typer av kontroller	21
3.2	Förändrad syn på kvalitetskontroll	22
3.3	Kontroll av lägesosäkerhet	24
3.4	Stickprov och kontrollområden	25
A	Exempel på kontrollmätningar och toleranser	28
A.1	Förenklad kvalitetskontroll	28
A.2	Kontroll av mätosäkerheten i Nätverks-RTK	29
A.3	Kontroll av utsättning	31
A.4	Kontroll av geoidmodell	33
A.5	Test av grova fel, systematik och mätosäkerhet i 1 D, 2D och 3D	34
A.6	Kontroll av grova fel, fullständighet, objekt- klassificering och logisk konsistens	37
A.7	Lägesosäkerhet – intervallskala	39
A.8	Lägesosäkerhet vs. antal värdesiffror	40
B	Specifikationer, checklistor m.m.	42
B.1	Att ta fram en dataproduktspecifikation	42
B.2	Att läsa en dataproduktspecifikation	47
B.3	Kontrollprocessen	48
B.4	Olika kontrollmetoders användbarhet	50
C	Ett tillämpningsexempel	51
C.1	Dataproduktspecifikationens krav	51
C.2	Test av tematisk osäkerhet	51
C.3	Test av fullständighet	52
C.4	Test av grova fel, systematik och lägesosäkerhet	52

D	Norskt register över kvalitetsmått	54
D.1	Inledning	54
D.2	Fullständighet.....	54
D.3	Logisk konsistens	54
D.4	Lägesosäkerhet	56
D.5	Tematisk osäkerhet	56
D.6	Temporal kvalitet.....	57
D.7	Användbarhet.....	58
D.8	Generellt kvalitetsmått.....	58
E	Statistik, introduktion	59
E.1	De viktigaste statistiska storheterna	59
E.2	Terminologi.....	59
E.3	Kontrollens trovärdighet	60
E.4	Kravet på signifikans.....	60
E.5	Konfidensintervall	60
E.6	Hypotesprövning.....	61
E.7	Kvalitetsredovisning vs. kvalitetskontroll	61

1 Inledning

Information

För eventuella fortlöpande justeringar av detta dokument, se [HMK-nytt](#).

I detta dokument ges en beskrivning av begreppet *geodatakvalitet*. Dessutom ingår ett antal metoder för kontroll av geodatakvalitet – från de enkla, till de mer strikta, ”statistiskt korrekta”. Dessa har samlats i bilagor – för att öka överblicken och underlätta metodvalet utifrån behov, ambitionsnivå med mera.

Formella standarder och andra influenser

HMK utgör en brygga mellan den formella standardiseringsverksamheten och det vardagliga arbetet för beställare, utförare och användare. Formella standarder är inte alltid samordnade och kompletta:

- De överlappar varandra, och det finns ibland motsägelser mellan dem.
- Det finns delar som inte täcks, eller delar som stannar vid det teoretiska och inte ger praktiska anvisningar
- Det finns behov av nationella, terminologiska anpassningar.

HMK kan också utgöra inkörsport till de formella standarderna. Dokumentet baseras i huvudsak på följande standarder:

- [SS-EN ISO 19115-1:2014](#), *Geographic information - Metadata - Part 1: Fundamentals*
Geografisk information – metadata – Del 1: Grunder
- [SS-EN ISO 19131:2008](#), *Geographic information – Data product specifications* (Engelska originalversionen utkom 2007; revidering pågår och ny version förväntas till 2019)
Geografisk information – Specifikation av datamängder
- [SS-EN ISO 19157:2013](#), *Geographic information - Data quality*
Geografisk information – Datakvalitet
- [SIS-ISO/TS 19158:2012](#), *Geographic information – Quality assurance of data supply*
Geografisk information – Kvalitetssäkring av dataförsörjning.

Information om dessa och övriga standarder i ISO:s 19100-serie om geografisk information återfinns på [SIS-TK323:s webbsida](#). Genom ett avtal mellan SIS (*Swedish Standards Institute*) och Lantmäteriet

finns det nu möjlighet att kostnadsfritt få tillgång till alla svenska geodatastandarder och -handböcker. Under en övergångsperiod utökas urvalet med cirka 30 internationella geodatastandarder. Aktuell information finns och ansökning kan göras på ovannämnda hemsida.

Den mest genomgripande förändringen i kvalitetsnomenklaturen jämfört med "gamla" HMK är introduktionen av GUM. De viktigaste skillnaderna är:

- *mätosäkerhet* i stället för *noggrannhet* som överordnad term
- *standardosäkerhet* i stället för *medelfel*
- *utvidgad mätosäkerhet* som benämning för uttryck som 2σ och 3σ ; tvåan respektive trean benämns *täckningsfaktor*.

Termen *fel* undviks i möjligaste mån. Det som tidigare benämndes *slumpmässiga* och *systematiska fel* ersätts med termerna *slumpmässiga* och *systematiska avvikelser* (eller *effekter*). Dock används fortfarande termen *grova fel*, eftersom det då är regelrätta felaktigheter.

I [HMK-Ordlista](#), senaste version, finns förklaringar till GUM-termer, fackuttryck och till de förkortningar som används. Där finns också en beskrivning av skillnaderna mellan [ISO 19157](#) och HMK:s GUM-influerade terminologi. Även [HMK-TR 2015:1](#) utgör ett komplement beträffande GUM och den geodatakvalitetsterminologi i övrigt som tillämpas i HMK.

Arbetet med geodatakvalitet i Sverige har bedrivits parallellt med motsvarande arbete i Norge. Framför allt har ett gemensamt synsätt vuxit fram, men det finns även exempel på direkt utbyte av text, tabeller etc. Exempel på delar där det senare har skett är framför allt avsnitt 3.3 samt bilagorna A.5-6, B.3-4 och D. De finns både i HMK-Geodatakvalitet och i den nya norska standarden *Geodatakvalitet*.

Georeferering och lägesosäkerhet

Underhand har behov framkommit av att närmare förklara två grundläggande termer - *georeferering* och *lägesosäkerhet* - samt att tydliggöra deras användning i HMK.

Georeferering innebär att knyta en geodatamängd till ett **officiellt referenssystem**. I Sverige innebär det i regel att koordinater och höjder anges i de nationella systemen *SWEREF99* respektive *RH2000*. I HMK är grundförutsättningen att geodata **är** georefererade, det ligger ju i själva begreppet "geo".

Georeferering ingår därför som en naturlig del vid geodatainsamling. Att **inte** georeferera – utan nöja sig med en lokal koordinat/höjdanslutning – är att betrakta som ett undantag från denna huvudregel, som endast tillämpas i speciella sammanhang. Exempel på tillämpningar med **lokala referenssystem** är:

- bygg- och anläggningsmätning i förhållande till icke-anslutna *byggplatsnät*
- dokumentation av byggnader, statyer etc. med hjälp av *terrester laserskanning*, där en redovisning i skannerns koordinatsystem ofta tillämpas.

Lägesosäkerhet är en utvidgning av GUM-termen mätosäkerhet. Läget beräknas ju normalt ur en serie mätningar, men för en **användare** av geodata är det osäkerheten i detta läge som är det intressanta, inte osäkerheten i de bakomliggande mätningarna.

För **producenten** däremot är det viktigt att ha kunskap om mätosäkerheten för att kunna leverera en slutprodukt med önskad lägesosäkerhet. Så båda termerna behövs parallellt.

Det går också att koppla ihop termerna georeferering och lägesosäkerhet:

- Osäkerheten i **georefererade** data benämns *absolut lägesosäkerhet*, dvs. osäkerheten i läget i förhållande till det officiella referenssystem som dessa geodata är anslutna till.
- Osäkerheten i geodata som är **lokalt anslutna** benämns på motsvarande sätt *lokal lägesosäkerhet*, t.ex. lägesosäkerheten **inom** en byggnad, en vägbana eller annan anläggning. (Sådana data **kan** vara anslutna till ett officiellt referenssystem, men då vanligen med en absolut lägesosäkerhet som är avsevärt större än den lokala osäkerheten.)

Till detta kommer:

- *relativ lägesosäkerhet*, som beskriver den **inbördes överensstämmelsen** (skillnaden) **mellan** två geodatamängder, t.ex. avvikelserna mellan två punktmoln och två stråk/körspår vid flyg- eller fordonsburen laserskanning.

Denna terminologi skiljer sig från den som tillämpas i de internationella standarderna inom ISO:s 19100-serie och i deras svenska motsvarigheter, se Tabell 1. Som synes är det mest en fråga om nyanser i språkbruket. HMK-terminologin ansluter dock även bättre till termen "georeferering", som inte förekommer i SS-ISO.

Absolut och lokal mätosäkerhet mäts vanligen i form av *standardosäkerhet i plan och höjd*; ett värde för varje punkt. De vanligaste måtten

på den relativa lägesosäkerheten är i stället *medelavvikelse* och *RMS*, dvs. en enda storhet får karaktärisera avvikelserna i samtliga punkter som ingår i jämförelsen.

Tabell 1. Några terminologiskillnader mellan HMK och ISO:s 19100-serie.

HMK	ISO 19100	
	Svenska	Engelska
mätosäkerhet	noggrannhet	accuracy
lägesosäkerhet	noggrannhet	accuracy
absolut lägesosäkerhet	absolut noggrannhet	absolute accuracy
lokal lägesosäkerhet	relativ noggrannhet	relative accuracy
relativ lägesosäkerhet	relativ noggrannhet	relative accuracy

Observera att en liten **absolut** lägesosäkerhet förutsätter en liten **lokal** osäkerhet, men inte tvärtom. Den **relativa** lägesosäkerheten kommer vanligen in i bilden som en kontrollmetod. Liten relativ osäkerhet – mellan de olika delar som beskriver ett geografiskt objekt – krävs både för att åstadkomma en liten lokal och en liten absolut lägesosäkerhet. Se nästa exempel som är en problemställning relaterad till [HMK – Fordonsburen laserskanning 2017](#).

Exempel: Fordonsburen laserskanning avser vanligen dokumentation av vägområden. Laserdata samlas in för ett *körspår* i taget, med olika körriktning och visst överlapp. Kontroll av den **relativa lägesosäkerheten** i höjd sker genom jämförelse mellan körspårens gemensamma delar och ett RMS-värde för avvikelserna beräknas som underlag för analysen.

Genom *stråkutjämnning* reduceras sedan den **lokala** lägesosäkerheten men för att säkerställa en liten **absolut** lägesosäkerhet krävs även inmätning av och anslutning till kända *kontrollpunkter/kontrollobjekt*. Projektets syfte avgör om det är den lokala eller den absoluta lägesosäkerheten som är viktigast, och som det därför bör ställas krav på. Alternativt kan krav ställas på både lokal och absolut osäkerhet, med strängare krav på det förstnämnda.

2 Datakvalitet

2.1 Definition av geodatakvalitet

I texten är kvalitetsterminologin preciserad genom att ersätta den allmänna termen *kvalitet* med *datakvalitet*, eller *geodatakvalitet*, när det avses. Det är dock ibland underförstått, och framgår av sammanhanget, vilken typ av kvalitet som avses. Avsteg från preciseringen har också gjorts för att terminologin ska överensstämma med, till exempel, officiella standarder.

En geodatamängds kvalitet redovisas med hjälp av ett antal *kvalitetsparametrar*, se avsnitt 2.7. Kvalitetsredovisningen kan göras på *detaljnivå* eller på en mer *övergripande nivå*. Den kan lagras i datamängden eller separat.

Ett av de främsta syftena med geografiska databaser är att ge möjlighet till sambearbetning och analys av data från olika databaser. Kvalitetsuppgifter kopplade till utgångsdata behövs för att avgöra möjliga analyser samt vilken datakvalitet slutprodukterna förväntas få. Behovet av en datakvalitetsredovisning accentueras ytterligare av att framtida användare och tillämpningar inte är kända i dag.

Standarden [ISO 19157](#) specificerar principerna för att beskriva datakvalitet för geografiska data. Det finns även andra typer av kvalitet som är aktuella vid geodattainsamling.

I [ISO 19158](#) definieras, förutom datakvalitet enligt [ISO 19157](#), även omfattning, tidplan och kostnad som kvalitetsparametrar vid geodattainsamling. Se [HMK-TR 2015:1](#) avsnitt 1.2.

2.2 Datakvalitetsmodell och informationsmodellering

Datakvalitetsuppgifterna används alltså för att bedöma om en viss datamängd uppfyller en viss användargrups behov. Några exempel på denna typ av uppgifter är:

- Vilka objekttyper innehåller datamängden och vilka egenskaper har objekten? Detta dokumenteras normalt i form av en *objekttypskatalog*.
- Hur bra är överensstämmelsen mellan datamängden och den verklighet den avses representera? Detta dokumenteras med hjälp av kvalitetsparametrarna.

Datakvalitetsmodellen ska beskriva de principer som ligger till grund för modelleringen av geodatakvalitet, till exempel vilka metadata som ska finnas.

Innehållet i geodata karaktäriseras ofta ur tre olika perspektiv: **Vad**, **var** respektive **när** – som svar på frågan: Vad skedde var och när? En geodatamängd är en förenkling av verkligheten, och innehåller bara det som anses viktigt eller av intresse. Därför kan olika datamängder över samma geografiska område vara olika.

Vilka delar av verkligheten som ska ingå i en datamängd, och hur de utvalda delarna ska representeras, beskrivs i en *informationsmodell*. Det är viktigt att denna modell är dokumenterad med hjälp av ett formellt *modelleringsspråk*, förslagsvis UML (*Unified Modeling Language*).

En informationsmodell innehåller:

- En beskrivning av de delar av "verkligheten" som ska vara med. I objekttypskatalogen beskrivs objekttyperna tillsammans med de egenskaper som ska ingå.
- De förenklingar som krävs för en effektiv representation.
- Vilka regler i datamängden som ska uppfyllas, till exempel topologiska egenskaper som att vägar ska bilda slutna nätverk.

För att kunna avgöra datakvaliteten för de olika delarna av en datamängd måste man alltså känna till både hur verkligheten är beskaffad och hur den är tänkt att representeras.

2.3 Datakvalitetsfilosofi

Begreppet kvalitet avser, enligt [ISO9000](#), bland annat en produkts förmåga att uppfylla användarbehoven. Datakvalitetsfrågorna är därför alltid viktiga i hanteringen av en geodatabas.

Databaser byggs upp och ajourhålls med olika metoder. Valet av metod sker med utgångspunkt från bland annat kvalitetskrav på slutprodukten, detaljeringsgrad på objekten och effektivitet. När valet är gjort sätts kraven på datakvaliteten för insamlingen/ajourhållningen utifrån denna metod.

Data insamlas för:

- *förstagångsuppbyggnad*; leverans av en homogen datamängd från ett insamlingstillfälle
- *ajourhållning* av en datamängd/databas över tiden.

Vad gäller det senare så skiljer vi dessutom på *kontinuerlig ajourhållning* ("omedelbar" uppdatering) och *periodisk ajourhållning* (återkommande uppdatering med jämna mellanrum). De mest typiska skillnaderna mellan förstagångsuppbyggnaden och ajourhållningen av en datamängd åskådliggörs i Tabell 2.3.

Tabell 2.3. Typiska skillnader mellan förstagångsuppbyggnad och ajourhållning av en datamängd – sett ur ett datakvalitetsperspektiv. (DPS = Dataproduktspecifikation, se avsnitt 2.5)

	Förstagångsuppbyggnad	Ajourhållning, kontinuerlig eller periodisk
Aktualitet	Lika för hela datamängden, men beror på underlaget	Beror på angivna ajourhållningsregler enligt DPS
Aktörer	Få aktörer, ofta ett upphandlat uppdrag	Flera aktörer inblandade
Insamlingsmetod	Enhetlig	Varierar med teknikutvecklingen
Krav	Enhetliga och tydliggjorda i DPS	Kan variera över tid; till exempel kan nya användningsområden tillkomma som ställer andra krav
Omfattning	En <i>omfattning</i> för hela datamängden	Datamängden indelas ofta i delomfattningar, till exempel geografiska, tematiska eller per objekttyp
Redovisning av datakvalitet	I enlighet med DPS; de enhetliga förhållandena gör att redovisningen ofta kan göras på övergripande nivå	Variationerna över tid innebär att redovisningen ibland måste göras på objektnivå
Typ av kontroll	Kontroll av hela datainsamlingsprocessen	Kontroll av slutresultatet av respektive ajourhållningsinsats
Underlag	Enhetligt	Varierar över tid; består delvis av objekt som redan finns i datamängden

Olika huvudmän befinner sig ofta i olika faser. Exempelvis är Trafikverkets investeringsverksamhet ofta inriktad mot **tidsbegränsade insatser lokalt** (brobyggnad, ombyggnad av väg och järnväg utefter en viss sträcka etcetera), vilket kopplar till den vänstra kolumnen. Verksamheten i Lantmäteriet, NVDB vid Trafikverket, kommuner med flera kännetecknas i stället av **ajourhållning inom ett regionalt/nationellt område**, som kopplar till den högra kolumnen.

Mycket av fokuset – i HMK totalt sett och i detta dokument – ligger på förstagångsinsamlingen. På sikt måste dock ajourhållningen hanteras utförligare – hur den bör utföras och redovisas.

I ajourhållningsskedet kommer likartade objekt att över tiden ha olika datakvalitet beroende på vilken metod och teknik som har använts vid varje uppdateringstillfälle. Till exempel kan geometrisk upplösning för bilder vid stereokartering variera över tiden. Olika geografiska områden kan dessutom ha olika metoder, varianter av metoder, olika ajourhållningsintervall och olika detaljeringsgrad för samma typ av objekt.

Som användare av geodata är det viktigt att ha tillgång till relevant information om enskilda objekts datakvalitet för att kunna bedöma datas användbarhet; om kompletteringar behöver göras eller om nykartering krävs.

2.4 Kvalitetsmärkning och metadata

Metadata betyder ”data om data”. De beskriver datamängdens innehåll, struktur och kvalitet. Syftet är att kunna söka, hitta och utvärdera data och tjänster. Tonvikten ligger på olika kvalitetsuppgifter. Metadata ska bland annat visa hur en datamängd kvalitetsmässigt överensstämmer med dataproduktspecifikationen (se nästa avsnitt).

I metadata redovisas därför även information om eventuella kvalitetsutvärderingar – men också om sådant som leverantör, tillgänglighet, restriktioner, rättigheter, ajourhållningsregler och pris.

Språkbruket har börjat divergera på senare tid, men i HMK definieras metadata på följande sätt:

- På *objektnivå* (märkning av varje enskilt objekt) betraktar vi kvalitetsmärkningen som en särskild sorts attribut, antingen kopplat till geometrin, eller till objektets övriga attribut.
- På *övergripande nivå* (per objekttyp, geografiskt tema etc.) används termen *metadata* för de kvalitetsuppgifter som finns, oavsett om dessa finns fysiskt i data eller ligger separat.

En reviderad internationell metadatastandard fastställdes 2014 ([SS-EN ISO/19115-1:2014](#)). Den tekniska rapporten [SIS-TR 14:2012 Geografisk information - Metadata på svenska](#) är en översättning och ett förtydligande av utvalda delar av den engelskspråkiga metadatastandard. Under SIS/TK489 finns också en rapport [Nationell metadataprofil - specifikation och vägledning](#) för att ta fram enhetliga metadata.

Se vidare avsnitt 2.13 och 3.6 i [HMK-TR 2015:1](#).

2.5 Dataproduktspecifikationen

En *dataproduktspecifikation* (DPS) är ett standardiserat sätt att definiera en dataprodukt. Den skrivs normalt innan någon datafångst överhuvudtaget har skett – för att presumtiva användare ska veta vad dessa data kommer att kunna användas till. Den lever sedan vidare med produkten. Dataproduktspecifikationer innehåller ett särskilt kvalitetskapitel och bör följa [SS-EN ISO 19131](#).

Se vidare Bilaga B.1 och B.2 samt avsnitt 2.12 i [HMK-TR 2015:1](#) och avsnitt 2.1 i [HMK – Introduktion 2017](#). Dataproduktspecifikation kallas även *dataspecifikation* av [Geodataportalen/geodata.se](#) och *geodata-specifikation* av [Svensk geoprocess](#).

2.6 HMK-standardnivåer

För att stödja geodatainsamling för några olika principiella användningsområden används begreppet *HMK-standardnivå*. HMK-standardnivå definieras som ”rekommendationer för beställarens val av metod och parametrar vid geodatainsamling för visst användningsområde”. Vald HMK-standardnivå är även styrande för utförarens genomförande i vissa delar.

HMK-standardnivåerna baseras dels på kraven från beställare, användare, tillämpningar och produkter – dels på tekniska produktionsmöjligheter. Standardnivåerna utgör bryggan mellan beställarens krav och utförarens teknikval och genomförande.

Fyra HMK-standardnivåer är definierade. Standardnivåerna ska hjälpa beställaren att hitta sin kravbild – där slutproduktens datakvalitet blir den önskade, oavsett utrustning, programvara etcetera. Nivåerna numreras från 0 och uppåt, där 0 är den nivå som har de lägsta kraven på datakvalitet. Det sker enligt följande:

0. Global/nationell mätning och kartläggning
1. Nationell/regional mätning och kartläggning
2. Mätning och kartläggning av tätort
3. Projektinriktad mätning och kartläggning

Denna konstruktion tillåter expansion i efterhand – om kvalitetskraven höjs eller teknikutvecklingen möjliggör kvalitetshöjningar.

HMK-standardnivå 0

HMK-standardnivå 0 används vid global/nationell mätning och kartläggning för dokumentation av markanvändning och vegetation, miljöövervakning med mera.

Kraven på lägesosäkerhet ligger på meternivå eller sämre. Normalt utförs geodatainsamling i HMK-standardnivå 0 med absolut positionering med GNSS och med bilddata från satelliter med en upplösning på halvmeternivå eller sämre. Vid presentation krävs ofta generalisering för att öka läsbarheten. Denna standardnivå hanteras inte inom ramen för HMK.

HMK-standardnivå 1

HMK-standardnivå 1 används för nationell/regional mätning och kartläggning för översiktlig planering och för dokumentation av byggande, infrastruktur, miljö, naturvård, risker, skogsbruk med mera.

Kraven på lägesosäkerhet ligger mellan meter- och decimeternivå. Normalt utförs geodatainsamling i HMK-standardnivå 1 med dGPS/dGNSS och med bilddata från flygplan med upplösning på halvmeternivå eller bättre. Vid presentation kan generalisering krävas för att öka läsbarheten.

HMK-standardnivå 2

HMK-standardnivå 2 används för mätning och kartläggning av tätort för kommunal detaljplanering och dokumentation.

Kraven på lägesosäkerhet ligger på decimeternivå eller bättre. Normalt utförs geodatainsamling i HMK-standardnivå 2 med Nätverks-RTK eller totalstation och med bilddata från flygplan med en upplösning på decimeternivå eller bättre.

HMK-standardnivå 3

HMK-standardnivå 3 används för projektinriktad mätning och kartläggning för projektering, byggande och förvaltning av bebyggelse, vägar och övrig infrastruktur samt bygg- och relationshandlingar.

Kraven på lägesosäkerhet ligger på 5-centimeternivå eller bättre. Normalt utförs geodatainsamling i HMK-standardnivå 3 med totalstation eller Nätverks-RTK (alternativt projektanpassad Nätverks-RTK) och med bilddata från helikopter eller markfordon, med en upplösning på halvdecimeternivå eller bättre.

Terrester laserskanning, förekommer också liksom gemensam insamling av bild- och laserdata från helikopter eller markfordon. Även bilddata insamlade med UAV¹ (*unmanned airborne vehicles*) börjar nu användas.

Sammanfattning

En sammanställning per HMK-standardnivå redovisas i Tabell 2.6.

¹ Benämns även UAS (Unmanned Airborne System), RPAS (*Remotely Piloted Aircraft System*) eller "drönare".

Tabell 2.6. Principiella exempel på ändamål och tekniska lösningar för HMK:s olika standardnivåer.

HMK-Standardnivå	0	1	2	3
Exempel på ändamål för mätning och kartläggning	Global/nationell bevakning av miljö och naturvård	Nationell/regional översiktlig planering	Detaljplanering av tätort	Projektering och byggande av infrastruktur och byggnader
Ungefärlig lägesosäkerhet	≥1m	≤1m	≤0,1 m	≤0,05m
Exempel på hantering av referenssystem	WGS84 ≈ SWEREF99	SWEREF 99 TM RH2000	SWEREF 99 regional projektionszon, RH2000	Projektanpassat system (inpassat på SWEREF99 och RH2000)
Exempel på geodetiska mätmetoder	Absolut positionering med GNSS	dGPS/dGNSS	RTK eller Nätverks-RTK	Totalstation, projektanpassad Nätverks-RTK och terrester laserskanning
Geometrisk upplösning vid mätning i bild-data	≥0,5m	≤0,5m	≤0,1m	≤0,05m
Principexempel avseende detaljeringsgrad för objektet byggnad	Redovisas som höjdsatt punkt	Redovisas som "låda"	Redovisas med enkel takkonstruktion och fasad utan detaljer	Redovisas med detaljerade takkonstruktioner, husliv och insida

2.7 Kvalitetsteman och kvalitetsparametrar

HMK:s *kvalitetsteman* för geodatakvalitet är *fullständighet, logisk konsistens, lägesosäkerhet, tematisk osäkerhet, temporal kvalitet samt användbarhet*. Detta följer [ISO 19157](#) sånär som på att i HMK används termen "osäkerhet" i stället för ISO-standardens "noggrannhet", se [HMK-Ordlista](#) (senaste version), kapitel 1.

Primära parametrar

Datakvaliteten beskrivs med hjälp av de *kvalitetsparametrar* som hör till respektive kvalitetstema, se Tabell 2.7. Till varje kvalitetsparameter kopplas sedan ett eller flera *kvalitetsmått* för att explicit "mäta" datakvaliteten.

Därutöver redovisas ibland *syfte, spårbarhet* samt *aktualitet*. Dessa kvalitetsteman finns omnämnda i standardiseringsarbetet men har inte formellt inkluderats i standarderna. Formen är av typen fri-text, så länge inte kvalitetsparametrar och kvalitetsmått har tagits fram.

Tabell 2.7. Översikt över kvalitetsteman och primära kvalitetsparametrar för datakvalitet. För detaljer och definitioner se [Bilaga D](#) i detta dokument samt [HMK - Ordlista](#), senaste version.

Fullständighet	Datamängdens innehållsmässiga överensstämmelse med dataproduktspecifikationen; brist eller övertalighet för objekt, attribut eller relationer.
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>brist</i> (för få) - <i>övertalighet</i> (för många).
Logisk konsistens	Överensstämmelse med logiska regler för datastruktur, attribut och relationer (till exempel sammanhängande nätverk och slutna ytor).
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>konceptuell konsistens</i> (följer det konceptuella/begreppsmässiga schemat) - <i>domänkonsistens</i> (giltig värde-mängd, dvs. tillåtna värden) - <i>formatkonsistens</i> (korrekt format) - <i>topologisk konsistens</i> (följer angivna topologiska regler).
Lägesosäkerhet	Osäkerhet i position.
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>absolut lägesosäkerhet</i> (i förhållande till referenssystemet) - <i>lokal lägesosäkerhet</i> (i förhållande till närliggande objekt, se Tabell 1) - <i>lägesosäkerhet hos rasterdata</i>.
Tematisk osäkerhet	Osäkerhet hos kvantitativa attribut och huruvida kvalitativa attribut och klassificeringar är korrekta.
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>klassificeringsosäkerhet</i> (korrekthet beträffande objekttyp) - <i>tematisk osäkerhet, kvalitativa attribut</i> (icke mätbara) - <i>tematisk osäkerhet, kvantitativa attribut</i> (mätbara).

Temporal kvalitet	Osäkerhet för temporala (tidsmässiga) attribut och temporala förhållanden mellan objekt.
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>tidsosäkerhet</i> (osäkerhet i tidsangivelse) - <i>temporal konsistens</i> (korrekthet i tidsordning) - <i>temporal validitet</i> (tidsuppgifters giltighet, till exempel att datum skrivs på angiven form).
Användbarhet	En datamängds lämplighet för en viss specifik tillämpning, som alltså styr bedömningen.
	Kvalitetsparametrar: Om övriga kvalitetsparametrar inte uttrycker en speciell datakvalitetsegenskap tillräckligt bra, kan användbarhet tillämpas. Den är fri till formatet.

Syfte

Att formulera ett tydligt *syfte* med en dataprodukt har många fördelar, för den som ansvarar för geodata, för den som producerar eller förädlar data och för användarna, till exempel:

- för att förstå hur data ska tolkas och ge avsedd effekt i en viss verksamhet
- för att tydliggöra den tänkta nyttan med data och tjänster som bygger på dessa data
- för att kunna utvärdera graden av överensstämmelse mellan användarens informationsbehov och tillgängliga data.

Ett tydligt syfte bör formuleras i dataproduktspecifikationen. Genom att definiera syftet kan man också revidera detta om en från början oförutsedd användning efterfrågas.

Spårbarhet

I en kvalitetsmedveten produktion av geodata är det viktigt att även redovisa datas *spårbarhet* bakåt i produktionskedjan; synonymer är *ursprung* eller *tillkomsthistorik*.

Spårbarheten är viktig av flera skäl, bland annat:

- En dataproducent kan, med kunskap om en datamängds tillkomsthistoria, spåra och åtgärda felkällor. Ett upptäckt fel leder då till att man utöver att korrigerar avvikande data även kan spåra orsaken och åtgärda den.
- Med kunskap om tillkomsthistoria och tidigare tillämpning kan en användare bedöma datamängdens användbarhet för ett visst syfte.

Data och information om spårbarheten bör lagras i databastillämpningens eller informationssystemets metadata.

Aktualitet

Aktualitet finns inte med som egen kvalitetsparameter i datakvalitetsstandardens [ISO 19157](#), även om det är en datakvalitetssegenskap som kunder ofta efterfrågar.

Aktualitet är ett sätt att i en producent-/leverantörsroll mäta och löpande redovisa hur data av olika typer åldras och förlorar i informationsvärde för avsedd tillämpning. Aktualiteten anges lämpligen som den tidpunkt då objektet senast, genom kontroll, konstaterades vara korrekt redovisat. Det räcker inte med att titta på det datum när data producerades.

Aktualitet kan också relateras till *ledtider för ajourhållning*. Långa ledtider, från det att en förändring har skett tills det att data är uppdaterade i databasen, vid till exempel kontinuerlig ajourhållning, medför att datamängden inte är helt aktuell.

2.8 Datakvalitet för bild- och laserdata

Standarden [ISO 19157](#) har en tonvikt på vektordata. Kvaliteten i t.ex. laserdata och bilddata hanteras inte på ett utförligt sätt. Kvalitetsmått för Lägesosäkerhet går dock att tillämpa, i vissa andra delar saknas såväl kvalitetsparametrar som kvalitetsmått.

Det mesta får plats i temat Användbarhet, även om det kan vara osmidigt. I kommande revisioner av HMK-Geodatakvalitet bör en lika distinkt form för datakvalitet för dessa geodatatyper tas fram.

Nedan redovisas ändå – som ett exempel – ett försök till strukturerad sammanställning av datakvalitet såsom den redovisas i HMK-Bilddata och HMK-Laserdata. Vissa delar är gemensamma medan andra är specifika för respektive datatyp.

Gemensamt

Redan angivande av HMK-standardnivå specificerar vissa kvalitetsparametrar – inte minst spårbarheten och datas användbarhet, men även lägesosäkerheten.

Lägesosäkerheten påverkas bl.a. av stödpunkter/stödytor: deras antal, placering, utformning och inmättingsosäkerhet. Detta kan i efterhand säkerställas via särskilda kontrollpunkter. Registrering och kontroll av tvärstråk utgör ett tänkbart komplement.

Laser- som bilddatainsamling sker i dag vanligen med GNSS/INS-stöd (kombinerad satellit- och tröghetsmätning). Därför måste krav även ställas på detta stödsystem.

Angivande av referenssystem i plan och höjd är en viktig användbarhetsaspekt och registreringstidpunkt anger datas aktualitet.

Samtidig insamling av bild- och laserdata kan ibland ge vissa kvalitets fördelar totalt sett, men inte alltid för specifika produkter av det ena eller andra slaget.

Specifikt för bilddata

HMK-standardnivå specificerar, för bilddata, valet av parametrarna ([HMK - Flygfotografering 2017](#), Tabell 2.3.1): geometrisk upplösning, lägesosäkerhet i plan och höjd, övertäckning, bildkvalitet, solvinkel/skugglängd och fotograferingsperiod.

Upplösningen påverkar direkt tolkningsmöjligheterna vid stereobearbetning och standardosäkerheten i plan och höjd måste anpassas till denna. Övertäckningen och kamerans öppningsvinkel påverkar insynen i bilderna och mätosäkerheten i höjd. Bildkvaliteten påverkar t.ex. möjligheterna att ta fram bra ortofoton. Solvinkel/skugglängd påverkar både stereobearbetning och ortofotoframställning och fotograferingsperioden slutproduktens användbarhet.

Val av flyghöjd och framtagning/kontroll av orienteringsdata ur blocktriangulering är också specifika bilddataaspekter på datakvaliteten. Fullständigheten säkerställs genom att kontrollera att det finns bilder och övertäckning över hela projektområdet.

Hela kontrollprocessen beskrivs i [HMK - Flygfotografering 2017](#), kapitel 4 och Bilaga C.

Specifikt för laserdata

HMK-standardnivå specificerar, för laserdata, valet av parametrarna ([HMK - Flygburen laserskanning 2017](#), Tabell 2.3.1): punkttäthet, standardosäkerhet i plan och höjd samt den maximala skanningsvinkeln.

Punkttätheten/punktavståndet är tillsammans med laserns träffyta på marken avgörande för hur små objekt som går att identifiera i ett laserpunktmoln. Skanningsvinkeln påverkar insynen mot markytan i skogs- och stadsmiljö.

Valet av HMK-standardnivå styr även – t.ex. vad gäller lägesosäkerheten – hur detaljerade markmodeller som kan tas fram (minsta grid-upplösning och lämplig ekvidistans för höjdkurvor).

Insamlingsperioden påverkar användbarheten. För exempelvis markmodell ger insamling mellan snösmältning och lövsprickning störst användbarhet (ingen snö, inga löv) men kan öka projektkostnaden.

Val av skanningsparametrar är en specifik laserdataaspekt på datakvaliteten. Fullständigheten säkerställs genom att kontrollera att hela projektområdet täcks av data, med stipulerad punkttäthet. Kontrollen av tematisk osäkerhet går ut på att säkerställa att markklassningen är korrekt, vilket är särskilt viktigt vid automatisk klassning. Hela kontrollprocessen beskrivs i [HMK - Flygburen laserskanning 2017](#), Kapitel 4 och Bilaga C.

3 Kontroll av geodata

För att säkerställa viss kvalitet enligt dataproduktspecifikationen och för att redovisa faktisk kvalitet för en geodatamängd måste data genomgå kvalitetskontroll.

3.1 Olika typer av kontroller

Direkta kontrollmetoder tillämpas på den datamängd som ska undersökas. Man skiljer på *interna* och *externa kontrolldata*,

Vid kontroll mot interna data sker kvalitetskontrollen mot data som finns i själva datamängden. Exempel på sådana kontrollmetoder är kontroll med programvara och 3D-visualisering.

Kontroll gentemot externa data sker mot kontrolldata utanför den undersökta datamängden. Exempel på sådana metoder är

- visuell kontroll mot ortofoto eller flygbilder
- kontroll mot andra datatyper, till exempel laserdata
- mätning i fält och annan fältkontroll.

Indirekta kontrollmetoder baseras på värdering av datakvaliteten med hjälp av andra källor än själva datamängden. Dessa källor kan exempelvis vara metadata, kunskap om produktionsmetoderna eller produktionsrapporter.

De indirekta metoderna innebär en subjektiv bedömning och ger begränsad möjlighet att kvantifiera datakvaliteten. Dessa metoder kan dock vara kostnadseffektiva och enklare att utföra – men resultatet kan inte fullt ut jämföras med dataproduktspecifikationens krav.

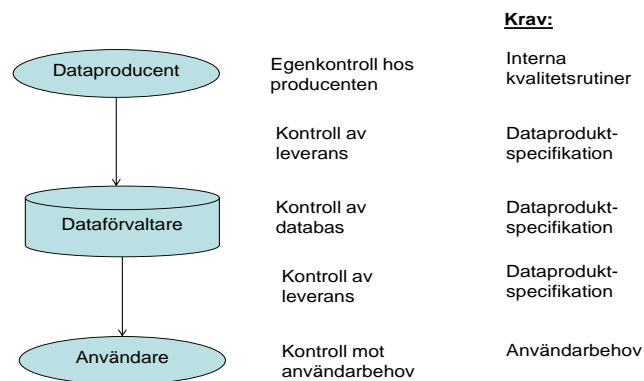
Med avseende på omfattningen skiljer vi på *delkontroll* och *full kontroll* (av hela datamängden).

Vid full kontroll kontrolleras samtliga förekomster, något som i regel kräver att kontrollen kan utföras maskinellt, till exempel ytbildning för att testa topologi (logisk konsistens). Full kontroll används också när antalet förekomster av en objekttyp är litet och stickprovskontroll inte tillämpbar.

Delkontrollen utförs efter tre principer: *stickprovskontroll*, *procentuell kontroll* och *statistisk kontroll*.

Egenkontroll avser den kontroll som utföraren av arbetet gör. Resultatet redovisas till beställaren, men denne kan även göra egna kontroller, *beställarens kontroll*.

I Figur 3.1 sätts olika typer av kontroll in i flödet
Producent → *Förvaltare* → *Användare*.

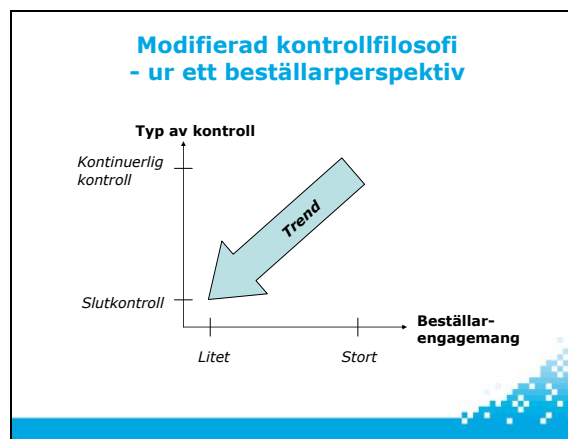


Figur 3.1. Olika typer av kontroller på vägen från producent till användare.

3.2 Förändrad syn på kvalitetskontroll

Synen på kontroll av datakvalitet håller på att förändras. Det beror bland annat på en övergång från verksamhet i egen regi till en beställar-/utförarmodell.

- De äldre HMK-skrifterna byggde på en detaljerad styrning av hela mätprocessen. Kraven utgick från en aktiv och mätningsskunnig beställare, samt checklistor och metodbeskrivningar med tillhörande toleranser. Slutprodukten kontrollerades endast i undantagsfall, till exempel vid detaljmätning.
- På senare tid har det skett en förskjutning mot att beställaren kontrollerar slutprodukten och mot att beställaren ger utföraren ett eget ansvar för kontroller under vägen (Figur 3.2). Det förekommer att beställarens slutkontroll upphandlas.



Figur 3.2. Utvecklingen vad gäller beställarkontroll av geodata går delvis mot slutkontroll och ett mer begränsat beställarengagemang i kontrollen av datafångstprocessen.

Förskjutningen har följande orsaker:

- Tekniken blir alltmer komplex och har inslag av "black-box"-karaktär, vilket gör det svårt att ge generella råd, anvisningar och toleranser för utförandet. Styrningen blir därför fabrikatberoende och inkluderad i den svarta lådan.
- Beställaren saknar ofta egen utförarkompetens och egna resurser för kontroll av utförandet.

Följande jämförelse är hämtad från laserskanning:

- Flygburen laserskanning innehåller i datafångsten flera moment som är fabrikatberoende. Leverantören tar eget ansvar för kontroller, men utfallet dokumenteras.
- För RTK-inmätning av stöd- och kontrollpunkter går det att sätta upp generella anvisningar och kontroller – vilket bör göras, inte minst som stöd- och kontrollpunkterna är centrala i hela laserdatahanteringen.
- Ett mellanläge har kontrollen av flygplanets positions- och attityd-/vinkelbestämning. Vissa delar är generella (till exempel *DOP-tal* och *cut-offvinkel*), medan andra beror på utrustningens fabrikat.

Fortfarande gäller att kontinuerliga kontroller under mätprocessens gång ger större möjligheter att åtgärda fel i ett tidigt skede. Även om det mer och mer blir utförarens ansvar att utforma och genomföra kontrollerna så bör naturligtvis kontrollresultatet redovisas. Se vidare [HMK - Introduktion 2017](#), avsnitt 2.2.

Ibland bör det nya synsättet väljas, i andra fall är det befogat att tillämpa synsättet i de äldre HMK-skrifterna. Ett överlåtande av kontrollen till leverantören ska ske under förarbetet genom att de två parterna – i en iterativ process – kommer fram till hur kontrollerna ska utföras och redovisas, vilket [ISO 19158](#) så väl beskriver.

Vid mindre uppdrag är en två-stegskontroll vanlig. Först genomförs en begränsad kontroll – om möjligt av slutprodukten – och endast om den uppvisar avvikelser sker en fördjupad kontroll. Det reducerar mängden kontroller, och därigenom kontrollkostnaden, då kontroller ger säkerhet i hanteringen men är improduktiva.

Vid mer omfattande projekt, eller vid längre beställar-/utförarförhållanden, kan det vara lämpligt att låta utföraren dokumentera sina delprocesser och utvärderingskriterier i en kvalitetsplan, som sedan testas med en provleverans. När provleveransen är godkänd av beställaren – eventuellt efter flera försök – ska/får leverantören inte ändra på produktionsprocessen. Detta ger beställaren kontroll över hela processen utan att detaljstyra den.

3.3 Kontroll av lägesosäkerhet

Toleranser

I HMK används numera begreppet *toleranser* i stället för det tidigare "felgränser".

Den överordnade termen är alltså:

- Tolerans, som antingen kan vara ett krav på den maximalt tillåtna avvikelser eller en specifikation av den variation som kan förväntas i ett visst sammanhang.

Inom mätningstekniken skiljs det på:

- *Produkttolerans*; maximal avvikelse från produktspecifikation.
- *Kontrolltolerans*; maximal avvikelse vid kontrollmätning.
- *Mättolerans*; maximal avvikelse från "sant värde" vid mätning; oftast av typen 2σ (dvs. 95 % signifikansnivå i enlighet med GUM).

Här hanteras mät- och kontrolltoleranser. Nödvändig mättolerans härleds från specificerad produkttolerans, men i den senare kan även andra felkällor i produktionsprocessen ingå, till exempel tillverkningsfel vid byggnation.

Grundläggande princip

Toleranserna för själva mätningen ska kopplas till den mätmetod som väljs (mättoleranser), inte till de krav som ställs på slutprodukten (produkttoleranser). Logiken är följande:

- Välj den metod som uppfyller produktkraven, och i övrigt är den optimala.
- Kontrollera mätningarna mot denna metods "inneboende" mätosäkerhet, det vill säga den mätosäkerhet som den valda metoden normalt ger.

Genom att följa denna logik går det att lita på metadata avseende till exempel Ursprung. Vi vet då att data märkta med inmätningssmetoden "Nätverks-RTK" eller "Totalstation" håller avsedd mätosäkerhet, eftersom kontrollerna har jämförts med den metod som faktiskt har använts. Om mätningarna kontrolleras mot toleranser som är generösa i förhållande till mätmetodens osäkerhet, kan besynnerliga effekter uppstå om grova fel förekommer.

Utförarna bör ha egna kvalitetssystem för hantering av instrument, programvaror och processer. Kontroller beskrivna i HMK är avsedda att komplettera de egna kvalitetssystemen samt peka på sådant som det är särskilt viktigt att ha under uppsikt.

Kontrollens genomförande

Geodatastandarder verkar ibland frångå beprövad praxis inom statistiken. Detta gäller främst lägesosäkerhet men även mätosäkerhet i de attributdata som mäts.

Bristerna består i huvudsak av följande:

- Kontrollerna är fokuserade på medelfel (dvs. det som i GUM benämns standardosäkerhet). Grova fel och systematiska avvikelser hanteras ej.
- Kontrollerna är väldigt enkla, för att inte säga förenklade. Man jämför till exempel beräknat medelfel med det teoretiska - rakt av och inte med ett statistiskt F-test.
- Kontrollerna tar ingen hänsyn till kontrollmetodens mätosäkerhet och stickprovets storlek.

Det hanterar såväl "gamla" HMK som den nya norska standarden Geodatakvalitet på ett mer korrekt sätt. Flera kompletteringar har därför gjorts i Bilaga A, för att åtgärda denna brist. Detta gäller såväl lägesosäkerheten som mätosäkerheten i kvantitativa attribut.

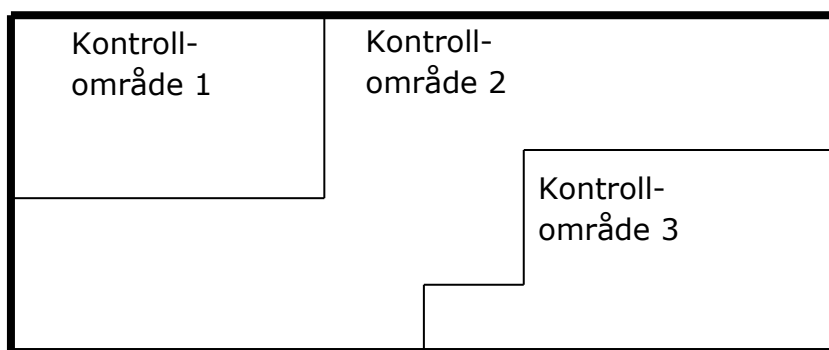
Systematisk avvikelse, standardavvikelse och RMS ger tillsammans en detaljerad beskrivning av lägesosäkerheten, se [HMK-TR 2015:1](#), avsnitt 2.10. I en sådan redovisning bör även ingå lokalisering av eventuella *grova fel*.

Beräkningsgången - exempelvis vid analys av en kontrollmätning av ett stickprov (se Bilaga A.2, A.5 och C.4) - blir:

- Eliminera ev. grova fel, antingen med någon speciell felsökningsmetod eller helt enkelt genom att tillämpa principen "avvikelser större än 3σ = grova fel".
- Beräkna systematisk avvikelse och standardavvikelse, eller RMS, utan de grova felen.
- Rapportera de grova felens antal, storlek, läge etcetera. Det ger information om vilken frekvens och storlek som kan förväntas vad gäller grova fel utanför stickprovet.

3.4 Stickprov och kontrollområden

Som utgångspunkt för stickprov delas datamängden lämpligen in i *kontrollområden* - helst på ett sådant sätt att de tillsammans täcker hela kartläggningområdet, se Figur 3.4.



Figur 3.4. En indelning i tre kontrollområden som tillsammans täcker kartläggningsområdet. (Fritt efter den norska standarden Geodatakvalitet)

Om till exempel kontrollområde 1 kontrolleras gäller utfallet för hela kontrollområde 1, men bara för detta. Det vill säga, om kontrollen överskrider givna toleranser i stickprovet så underkänns hela data-mängden i det kontrollområdet. Hur en sådan brist åtgärdas är något som måste avtalas mellan beställare och utförare. Minst två objekttyper bör kontrolleras.

Tabell 3.4. Stickprovsstorlekar (minimalt antal objekt) för kvalitativ och kvantitativ kvalitetskontroll. (Identisk med den norska standarden Geodatakvalitet)

Antal förekomster av objekttyper i kontrollområdet		Stickprovsstorlek (kvalitativ kontroll)	Stickprovsstorlek (kvantitativ kontroll)
Från	Till		
1	8 (5)	Alla objekt	Alla objekt
9 (6)	50	8	5
51	90	13	7
91	150	20	10
151	280	32	15
281	400	50	20
401	500	60	25
501	1200	80	35
1201	3200	125	50
3201	10 000	200	75
10 001	35 000	315	100
35 001	150 000	500	150
150 001	500 000	800	200
> 500 000		1250	200

Stickprovets storlek beror på hur många objekt av den aktuella objekttypen som finns i datamängden, men till del också på vad som ska kontrolleras.

Där företeelsen är *kvalitativ*, till exempel test av grova fel eller fullständighet (finns/finns inte), krävs större stickprov än vid exempelvis kontroll av standardavvikelse eller systematik. I det senare fallet bedöms man få mer information eftersom avvikelserna ligger på en *kvantitativ*, steglös intervallskala.

En guide för att bestämma minsta stickprovsstorlek för olika antal objekt i ett kontrollområde redovisas i Tabell 3.4. Se vidare tillämpningsexemplet i Bilaga C.1.

A Exempel på kontrollmätningar och toleranser

Här redovisas några exempel på hur kontrollmätningar kan hanteras. Beskrivningen går från de enklaste till de mer teoretiskt stringenta metoderna. Dessa bygger på etablerade analysförfaranden, som dock inte redovisas i detalj.

Kontrollinsatsen ska stå i rimlig relation till kostnaden för insatsen och till uppdragets omfattning, så här finns möjlighet att välja metod utifrån vars och ens behov.

- A.1 behandlar de enkla/förenklade kontrollmetoderna.
- I A.2 studeras mätosäkerheten i nätverks-RTK, inklusive sökning av grova fel och analys av systematik. Det är en enkel tillämpning av HMK:s trenivåmetod (1σ , 2σ , 3σ) som kan användas även i andra osäkerhetsanalyser.
- A.3 hanterar kontroll av utsättning och A.4 kontroll av en geoidmodell. De är båda generella och baseras på traditionell variansanalys. Formlerna för beräkning i Excel anges.
- A.5 utgör en mer utvecklad metod för analys av grova fel, standardosäkerhet och systematik än den som redovisas i A.2. Bland annat görs en distinktion mellan 1D, 2D och 3D.
- I A.6 redovisas en metod för att mer nyanserat bedöma kontroller av grova fel och fullständighet/objektklassificering ur stickprov. Den innebär att vissa sådana avvikelser får förekomma och reglerar hur många som får finnas i förhållande till stickprovets storlek.

A.1 Förenklad kvalitetskontroll

Förenklade kontrollmetoder är sådana där man jämför till exempel beräknat medelfel med det teoretiska – rakt av och inte med ett statistiskt F-test (se avsnitt 3.3). Kontrollerna tar heller ingen hänsyn till kontrollmetodens mätosäkerhet och stickprovets storlek.

Användningen kan dock vara väl motiverad i många sammanhang. I boken [Geografisk informationsbehandling](#) (Harrie et.al., 2013, kap 10) ges en bra redovisning av sådan kvalitetskontroll.

Vid användning av de förenklade kontrollmetoderna sätter man alltså *dataproduktspecifikationens datakvalitetskrav* lika med *toleransen* vid *kvalitetskontroll*, det vill säga

$$\text{krav} = \text{tolerans}$$

Detta strider mot den traditionella statistikens krav på *signifikans*: Man måste lägga på en "säkerhetsmarginal" för att säkerställa att ett överskridande av toleransen ska betraktas som en **signifikant** avvi-

kelse från kravet (se avsnitt E.4). Vid kontroll av geodata – till exempel vad gäller lägesosäkerhet, antal grova fel, felklassificerade objekt etcetera – så innebär säkerhetsmarginalen normalt att toleransen är större/generösare än det specificerade kravet, dvs.

$$\text{krav} < \text{tolerans}$$

Att sätta kravet och toleransen lika innebär att något av följande två problem måste hanteras:

- Om vi antar att det gemensamma värdet (krav = tolerans) ska tolkas som **tolerans**, och om man klarar den, så är den egentliga datakvaliteten bättre (lägre felfrekvens, mindre mätosäkerhet etcetera) än detta värde.
- Om däremot det gemensamma värdet ska tolkas som **specificerad datakvalitet** så finns det en risk att data som uppfyller kraven ändå förkastas.

Exempel A.1: I en studie av Nationell höjdmodell (HMK-TR 2014:1) visas att standardosäkerheten i höjd är ca. 50 mm. För att lägesosäkerheten i ett skanningsområde ska anses **signifikant** avvika från detta värde – med sannolikheten 95 % – måste avvikelserna (RMS) i kontrolltytorna i genomsnitt överskrida 80 mm, dvs.

$$\text{specificerat krav: } 50 \text{ mm} \Rightarrow \text{tolerans: } 80 \text{ mm}$$

Att kräva att alla RMS-värden är ≤ 50 mm skulle förkasta cirka hälften av skanningsområdena (46 % i nämnda studie), i onödan. Medelvärde av alla kontrolltytors RMS är just 50 mm. Och att säga att standardosäkerheten i Nationell höjdmodell i höjd inte är bättre än 80 mm vore att utesluta många tillämpningar – också det i onödan eftersom det korrekta värdet alltså är 50 mm. Slutsats: Skilj på krav och tolerans!

A.2 Kontroll av mätosäkerheten i Nätverks-RTK

Följande exempel avser kontrollmätning av kända punkter för att verifiera eller uppskatta mätosäkerheten i Nätverks-RTK.

Minst 20 kontrollmätningar bör göras. Finns inte 20 stycken kända punkter bör varje punkt besökas flera gånger (med någon timmes mellanrum) för att uppnå ett 20-tal mätningar. Återbesök rekommenderas även av andra skäl, till exempel att mätningarna då blir mer okorrelerade.

Ett tillvägagångssätt är att kontrollera om en viss standardosäkerhet (σ) uppfylls genom att använda följande formler, som överensstämmer med avsnitt 3.3 (n = antalet kontrollmätningar, ε_i är avvikelserna från känt värde, l_i är en enskild kontrollmätning och \bar{l} är medelvärdet av samtliga sådana mätningar):

- Ingen avvikelse större än 3σ .
- Max en avvikelse större än 2σ (1 = 5 % av 20).
- 2/3 av mätningarna med en avvikelse inom $\pm 1\sigma$.
- Absolut medelavvikelse $|\bar{\varepsilon}| = \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \right|$ mindre än $\frac{2\sigma}{\sqrt{n}}$.
- Beräknad standardosäkerhet $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (l_i - \bar{l})^2}$ mindre än $\sigma \cdot (0,96 + n^{-0.4})$ (baserat på en F-fördelning och 95 % signifikansnivå, se [HMK-Stommätning](#), sid. 89).

Om till exempel $\sigma = 25$ millimeter (Nätverks-RTK i höjd) och 20 kontrollpunkter används blir toleranserna:

- Ingen avvikelse större än $3\sigma = 75$ mm.
- Max en avvikelse större än $2\sigma = 50$ mm.
- 13 mätningar med en avvikelse inom ± 25 mm.
- Absolut medelavvikelse $|\bar{\varepsilon}| \leq \frac{2\sigma}{\sqrt{n}} = \frac{2 \cdot 25}{\sqrt{20}} = 11$ mm.
- Beräknad standardosäkerhet mindre än $\sigma \cdot (0,96 + n^{-0.4}) = 25 \cdot (0,96 + 20^{-0.4}) = 32$ mm.

Toleranserna i plan, med $\sigma = 15$ mm, blir:

- Ingen (radiell) avvikelse större än $3\sigma = 45$ mm.
- Max en (radiell) avvikelse större än $2\sigma = 30$ mm.
- 13 mätningar med en (radiell) avvikelse inom ± 15 mm.
- Radiell medelavvikelse ² $\bar{\Delta}_R = \sqrt{\bar{\varepsilon}_N^2 + \bar{\varepsilon}_E^2} \leq \frac{2\sigma}{\sqrt{n}} = \frac{2 \cdot 15}{\sqrt{20}} = 7$ mm.
- Beräknad standardosäkerhet mindre än $\sigma \cdot (0,96 + n^{-0.4}) = 15 \cdot (0,96 + 20^{-0.4}) = 19$ mm.

Även om inga toleranser har angetts så bör ett osäkerhetsmått beräknas, till exempel mätningarnas standardosäkerhet. Då kan ett konfidensintervall runt skattningen göras för att även ange skattningens osäkerhet. En signifikansnivå på 95 % ger de täckningsfaktorer som redovisas i Tabell A.2.

² Först beräknas alltså de genomsnittliga avvikelserna (medelvärden) för Northing (N) resp. Easting (E). Därefter beräknas det radiella värdet $\bar{\Delta}_R = \sqrt{\bar{\varepsilon}_N^2 + \bar{\varepsilon}_E^2}$. Se vidare bilagorna A.5 och C.4.

Tabell A.2. Täckningsfaktorer för konstruktion av 95 %-iga konfidensintervall för standardosäkerhet.

Antal kontrollpunkter	Undre gräns $k_{low} \cdot \sigma$	Övre gräns $k_{high} \cdot \sigma$
1	0,45	31,91
2	0,52	6,28
3	0,57	3,73
4	0,60	2,87
5	0,62	2,45
7	0,66	2,04
10	0,70	1,76
15	0,74	1,55
20	0,77	1,44
30	0,80	1,34
40	0,82	1,28
50	0,84	1,24

Exempel A.2: Om standardosäkerheten har skattats till 15 millimeter blir konfidensintervallet för 5 kontrollpunkter:

$$[0,62 \cdot 15; 2,45 \cdot 15] = [9,3; 36,8]$$

det vill säga, det sanna värdet ligger med 95 % sannolikhet mellan 9,3 och 36,8 millimeter, vilket är ganska intetsägande.

Intervallet för 20 kontrollpunkter blir:

$$[0,77 \cdot 15; 1,44 \cdot 15] = [11,6; 21,6]$$

som är snävare.

Vid 50 punkter blir resultatet:

$$[0,84 \cdot 15; 1,24 \cdot 15] = [12,6; 18,6]$$

som inte är mycket mer exakt än resultatet för 20 punkter. Därav valet av minst 20 kontrollpunkter i detta avsnitt.

A.3 Kontroll av utsättning

En utsättning av 100 höjder ska kontrolleras med uttag av 20 slumpmässiga stickprov för kontrollmätning. Denna mätning ger de avvikelser i millimeter (kontrollmätning – produktionsmätning)³ som redovisas i Tabell A.3.a.

³ Denna teckenkonvention ger direkt rätt tecken på den korrektur vi räknar fram. Ett fel beräknas tvärtom (produktionsmätning – kontrollmätning)

Tabell A.3.a. Kontroll av utsättning. Avvikelse mellan kontrollmätning och ursprungsmätning. Enhet: millimeter.

Nr 1-5	Nr 6-10	Nr 11-15	Nr 16-20
14	7	-11	2
3	8	1	-1
2	-2	7	-19
16	2	26	-4
-8	-12	-7	-18

Kontroll ska göras av att utsättningen uppfyller kraven på 1, 2 respektive 3σ enligt HMK samt att standardosäkerheten (σ) inte överskrider dataproduktspecifikationens angivna krav på 10 millimeter. Kontrollmetoden kan för enkelhets skull betraktas som felfri.

Lösning

Avvikelse rangordnas och indelas i kategorier, se Tabell A.3.b.

Tabell A.3.b. Kontroll av utsättning. Sortering av avvikelser.

Intervall	Avvikelse i intervallet	Antal
Avvikelse $\leq 1\sigma$	3, 2, -8, 7, 8, -2, 2, 1, 7, -7, 2, -1, -4	13 st.
$1\sigma < \text{Avvikelse} \leq 2\sigma$	14, 16, -12, -11, -19, -18	6 st.
$2\sigma < \text{Avvikelse} \leq 3\sigma$	26	1 st.
Avvikelse $> 3\sigma$	-	0 st.

Utfallet jämförs med HMK i Tabell A.3.c.

Tabell A.3.c. Kontroll av utsättning; jämförelse med HMK:s gränsvärden.

	Normalfördelningens gränsvärde (%)	Kontrollresultat (%)
Avvikelse $\leq 1\sigma$	67	$13/20 \approx 65$
Avvikelse $\leq 2\sigma$	95	$(13+6)/20 \approx 95$
Avvikelse $\leq 3\sigma$	100	$(13+6+1)/20 \approx 100$
Avvikelse $> 3\sigma$	0	0

Resultatet har bra överensstämmelse med HMK:s gränsvärden. Dock bör kontrollmätning nummer 14 (26 millimeters avvikelse) göras om, eftersom den överskrider varningsgränsen 2σ . Om resultatet förblir detsamma kan beräkningen fortsätta.

Antag att mätningen kan godkännas, det vill säga att inga grova fel finns i materialet. Eftersom kontrollmetoden antas vara felfri beräknas då den kvadratiske medelavvikelsen (RMS), för att jämföra med den specificerade standardosäkerheten 10 millimeter.

Mätvärdena förs in i ett Excel-ark och följande beräkning görs:

$$\text{RMS} = \sqrt{\frac{\sum \text{avvikelser}^2}{\text{antal}}} \text{ som i Excel skrivs}$$

$$\text{RMS} = \text{ROT}(\text{KVADRATSUMMA}(A1:A20)/20) = 10,9 \text{ mm}$$

Toleransen/gränsvärdet för RMS blir

$$T_{\text{RMS}} = \sigma \cdot (0,96 + n^{-0,4}) = 10 \cdot (0,96 + 20^{-0,4}) \approx 10 \cdot 1,26 = 12,6 \text{ mm}$$

där σ är specificerad standardosäkerhet och n = antalet kontrollpunkter (här 10 millimeter respektive 20 stycken). Det innebär att det beräknade RMS (10,9) klarar toleransen för standardosäkerheten (12,6).

Sammantaget kan det konstateras att mätosäkerheten är i enlighet med specifikationen samt att inga grova fel har hittats i stickprovet; även mätning nummer 14 ligger inom vad som får betraktas som normala variationer i ett normalfördelat material (endast 1 på 20 överskrider 2σ , dvs. 5 %). Utsättningen bör därför godkännas.

A.4 Kontroll av geoidmodell

Geoidmodellen SWEN08_RH2000 ⁴⁾ ska kontrolleras i ett lokalt område genom jämförelse mellan Nätverks-RTK-bestämda och avvägda höjder på 20 regelbundet fördelade punkter i området. Mätningarna gav följande resultat (avvägning - RTK, enhet millimeter), se Tabell A.4.a.

Tabell A.4.a. Kontroll av geoidmodell. Avvikelse mellan kontrollmätning och ursprungsmätning. Enhet: millimeter.

Nr 1-5	Nr 6-10	Nr 11-15	Nr 16-20
53	39	3	29
31	41	27	23
29	21	39	-13
57	29	65	17
9	1	11	-11

Finns det något signifikant höjdskick, det vill säga har geoidmodellen någon märkbar systematisk avvikelse från höjdsystemet i detta område?

Lösning

Mätvärdena förs in i ett Excel-ark och RMS, standardosäkerheten (s) samt medelvärdet (m) beräknas, se Tabell A.4.b.

⁴⁾ En ny geoidmodell är under utveckling men har ännu inte lanserats officiellt

Tabell A.4.b. Kontroll av geoidmodell. Beräkning av RMS, standardavvikelse och medelvärde.

Storhet	Excelformel	Värde (mm)
RMS	ROT(KVADRATSUMMA(A1:A20)/20)	32,39
s	STDEVA(A1:A20)	21,13
m	MEDEL(A1:A20)	25,00

Standardosäkerheten (s) är spridningen kring medelvärdet. RMS är spridningen kring det "sanna" värdet - inklusive eventuell okänd systematik - om avvägningen betraktas som felfri i sammanhanget. Därmed är ett litet s-värde i förhållande till RMS en indikation på en systematisk avvikelse, ett så kallat *skift*.

Höjdavvikelsen kan anses vara signifikant, på 95 % nivå, om:

$$s/\text{RMS} \leq 1 - (n-1)^{-0.4} = 1 - 19^{-0.4} \approx 0,69$$

I det här exemplet är $s/\text{RMS} = 0,65$, vilket innebär att höjdskitet är signifikant och kan med hjälp av beräknat medelvärde skattas till +25 millimeter. Beräkningen ger även en skattning av standardosäkerheten i höjdbestämningen: 21 millimeter. Med antagandet om felfri avvägning är detta en skattning av RTK-mätningens standardosäkerhet.

Till de aktuella RTK-mätvärdena bör alltså 25 millimeter adderas för att överensstämna med höjdsystemet i området. Om syftet är att bestämma ett generellt höjdskitet bör urvalet utvidgas till fler stickprov. Det kanske krävs både en och två upprepningar innan ett säkert värde kan åstadkommas. Säkerheten beror på spridningen mellan de upprepade bestämningarna av skiftet.

A.5 Test av grova fel, systematik och mätosäkerhet i 1 D, 2D och 3D

Här förs ett resonemang om att införa en nyansering av gränsvärdena i en dimension, två dimensioner (i planet) och tre dimensioner (i rummet) bör införas. Detta för att få en mer teoretiskt korrekt signifikansnivå än det förenklade synsättet i Bilaga A.2.

Flera av de angivna värdena har verifierats genom Monte Carlo-simulering ([HMK-TR 2013:1](#)).

Grova fel

3σ (där σ är den **specificerade** standardosäkerheten) används som gräns för grova fel. I 1D motsvarar det signifikansnivån 99,73 %. Denna signifikansnivå i två och tre dimensioner ger $2,43\sigma$ i 2D och

2,17 σ i 3D, se Tabell A.5.a. Vill man söka grova fel med samma risknivå så är det alltså dessa täckningsfaktorer som ska användas.

Tabell A.5.a. Den gräns för grova fel i 2D och 3D som motsvaras av 3 σ i 1D.

1D	3 σ	→	99,73 %
2D	2,43 σ	←	99,73 %
3D	2,17 σ	←	99,73 %

Detta förutsätter dock att koordinatosäkerheterna σ_N , σ_E respektive σ_H (för *Northing*, *Easting*, *Höjd*) är någorlunda lika stora samt att det inte finns några större korrelationer mellan koordinatvärdena.

Är inte dessa villkor uppfyllda så tenderar täckningsfaktorerna att gå mot 3, dvs. gränsvärdena i 1D-raden ger **minst** 95 % signifikansnivå även i 2D och 3D. Så vill man hålla risknivån (sannolikheten för att underkänna en korrekt mätning) under 5 % så är det den kolumnen man alltid bör använda. Så är till exempel HMK:s traditionella 3-nivåprincip (1, 2 respektive 3 σ) upplagd – mest på grund av sin enkelhet, se avsnitt 3.3.

Mätosäkerhet

Den genomsnittliga mätosäkerheten – eller lägesosäkerheten – kontrolleras med hjälp av gränsvärden/toleranser för **beräknad** standardosäkerhet. Dessa beräknas ur F-fördelningen och även här finns en möjlighet att göra en nyansering av gränsvärdena beroende på dimensionen.

Nyanseringen sker genom att man ersätter frihetsgraderna $n-1$ med $dim(n-1)$, där dim (=dimension) antar värdena 1, 2 och 3 för 1D, 2D respektive 3D. Resultatet redovisas i Tabell A.5.b. Exempel på tabellens användning ges i Bilaga C.4.

Minskningen av gränsvärdena för 2D och 3D beror på att sannolikheten för stora koordinatosäkerheter minskar ju fler avvikelser som ingår; att till exempel *Northing*, *Easting* och *Höjd* **samtidigt** ska anta stora värden är liten. Förutsättningarna är desamma som vid detektering av grova fel, dvs. ungefär lika stora koordinatosäkerheter och inga nämnvärda korrelationer. Vill man gardera sig mot detta använder man 1D-kolumnen även i 2D och 3D.

Tabell A.5.b. Gränsvärden för test av standardosäkerheter i 1D, 2D och 3D. Signifikansnivå 95 %. (Identisk med den norska standarden Geodatakvalitet)

n	1D $\sqrt{F_{0.05,n-1,\infty}}$	2D $\sqrt{F_{0.05,2(n-1),\infty}}$	3D $\sqrt{F_{0.05,3(n-1),\infty}}$
5	1,54	1,39	1,32
7	1,45	1,32	1,27
10	1,37	1,27	1,22
15	1,30	1,22	1,18
20	1,26	1,19	1,15
25	1,23	1,17	1,14
35	1,20	1,14	1,11
50	1,16	1,12	1,10
75	1,13	1,09	1,08
100	1,12	1,08	1,07
150	1,09	1,07	1,05
200	1,08	1,06	1,05

Systematik

Eventuell systematik kontrolleras via gränsvärden för absolut (1D) eller radiell medelavvikelse (2D och 3D), jämför Bilaga A.2. Dessa beräknas ur t-fördelningen och varianter av denna. Även här finns en möjlighet att göra en nyansering. Istället för en strikt formel så har vi gått den enkla vägen och tillämpat Monte Carlo-simulering (100 000 simuleringar). Det ger det resultat som redovisas i Tabell A.5.c.

Tabell A.5.c. Gränsvärden för test av systematik i 1D, 2D och 3D. Signifikansnivå 95 %. (Identisk med den norska standarden Geodatakvalitet)

n	1D $t_{0,975,n-1}$	2D empiriskt t-värde	3D empiriskt t-värde
5	2,78	2,11	1,86
7	2,45	1,97	1,78
10	2,26	1,88	1,72
15	2,14	1,83	1,68
20	2,09	1,80	1,66
25	2,06	1,78	1,65
35	2,03	1,77	1,64
50	2,01	1,76	1,64
75	1,99	1,75	1,63
100	1,98	1,74	1,62
150	1,98	1,74	1,62
200	1,97	1,74	1,61
∞	1,96	1,73	1,61

Ett exempel på tabellens användning ges i Bilaga C.4.

Förutsättningarna är desamma som vid test av grova fel och mätosäkerhet, dvs. om man vill gardera sig mot inhomogen mätosäkerhet och/eller korrelation mellan koordinatvärdena så använder man 1D-kolumnen även i 2D och 3D.

A.6 Kontroll av grova fel, fullständighet, objektklassificering och logisk konsistens

Tabell A.6 gäller för kontroll av storheter som antingen är korrekta eller felaktiga. Dvs. de är *binära*, "noll" eller "ett", eller *logiska*, true/false; det finns inget däremellan. De måste kontrolleras på ett annat sätt än storheter som mäts på en skala, t.ex. *mätosäkerhet* som mäts "steglöst" med hjälp av standardosäkerheten. Exempel är kontroll av:

- antalet grova fel (lägesosäkerhet)
- antalet saknade eller övertaliga objekt (fullständighet)
- antalet felklassificerade objekt (tematisk osäkerhet)
- antalet formatfel i data (logisk konsistens).

Kontrollen sker mot stickprov och går ut på att jämföra antalet hittade felaktigheter med tabellvärdena för antalet tillåtna fel för den aktuella stickprovsstorleken. De gulmarkerade siffrorna i tabellen är subjektivt satta medan övriga erhålls ur formeln för *binomialfördelningens kumulativa fördelningsfunktion*

$$F(k) = P(X \leq k) = \sum_{i=0}^k \binom{n}{i} \cdot (1-p_o)^i (p_o)^{n-i} \quad (1)$$

där $P(\square)$ betecknar sannolikhet och X betecknar ett "lyckat utfall". p_o är den andel felaktigheter som accepteras i dataproduktspecifikationen, dvs. $P(X) = 1 - p_o$

Först bestäms storheten k_o som det minsta heltal (en integer $\leq n$) för vilken

$$F(k_o) = P(X \leq k_o) \geq 5\% \quad (2)$$

dvs.

$$F(k_o - 1) = P(X \leq k_o - 1) < 5\% \quad (3)$$

Det innebär att sannolikheten för att få $n - (k_o - 1)$ stycken felaktigheter är mindre än 5 %. Det vill säga toleransen blir

$$\text{tolerans} = n - (k_o - 1) = n - k_o + 1 \quad (4)$$

I programmeringsspråket MATLAB skrivs detta

$$\text{tolerans} = n - \text{binoinv}(0.05, n, 1 - p_o) + 1 \quad (5)$$

där *binoinv* står för "Binomial inverse cumulative distribution function". Ett alternativ till Tabell A.6 är alltså att producera gränsvärdena med hjälp av ett enkelt datorprogram.

Eftersom man har valt att antalet felaktigheter ska vara **mindre** än tabellvärdena blir i realiteten kraven strängare än binomialfördelningens teoretiska värden för angivna signifikansnivåer.

Tabell A.6. Kontroll av grova fel, fullständighet, objektklassificering och logisk konsistens ur stickprov. Antalet felaktigheter ska vara mindre än tabellvärdena. (Identisk med den norska standarden Geodatakvalitet)

Antal förekomster av objekttypen i kontrollområdet		Stickprovstorlek (n)	$p_o =$ andel tillåtna fel enligt dataproduktspecifikationen					
			0.5%	1.0%	2.0%	3.0%	4.0%	5.0%
Från	Till		Förkastningsgräns/tolerans					
1	8	Alla objekt	1	1	1	1	1	1
9	50	8	1	1	1	2	2	2
51	90	13	1	1	2	2	2	3
91	150	20	1	2	2	3	3	4
151	280	32	1	2	3	3	4	4
281	400	50	2	3	3	4	5	6
401	500	60	2	3	4	5	6	7
501	1200	80	3	3	5	6	7	8
1201	3200	125	3	4	6	8	10	11
3201	10 000	200	4	6	8	11	14	16
10 001	35 000	315	5	7	12	16	20	23
35 001	150 000	500	6	10	16	23	28	34
150 001	500 000	800	9	14	24	33	42	51
> 500 000		1250	12	20	34	49	63	76

Tabell A.6 utgör en tillämpning av Tabell 3.4 för kvalitativ kontroll. Ett exempel på deras tillämpning finns i Bilaga C.1.

A.7 Lägesosäkerhet – intervallskala

Lägesosäkerheten i geometriska data redovisas primärt i form av standardosäkerhet. För punktobjekt avser uppgifterna den enskilda punktens mätosäkerhet – för mer komplicerade objekt de ingående koordinaternas genomsnittliga osäkerhet.

Standardosäkerheten redovisas separat i plan och höjd. I skattningen ingår enbart osäkerheten i lägesbestämningen – inte den osäkerhet som beror på att skilda objekt (objekttyper) är olika lätta att definiera, eller att man har valt en förenklad geometrisk representation.

Det finns ett behov av en gemensam intervallhantering för lägesosäkerhet. Annars är risken stor att osäkerhetsuppgifterna spretar och blir oöverskådliga, samt att det blir svårt att överföra sådan information om datakvaliteten mellan olika system.

HMK förordar en intervallindelning för standardosäkerhet, med gränserna:

10 mm, 20 mm, 50 mm, 100 mm, 200 mm, 500 mm, 1000 mm.

Skalan kan expanderas både uppåt och nedåt till:

... 1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000 mm ...

Exempel A.7: I Stockholm stad har man definierat följande krav på geodetisk detaljmätning för olika produkter (standardosäkerhet):

- **Baskarta;** plan 50 mm, höjd 50 mm (hårdgjorda ytor) annars 100 mm.
- **Nybyggnadskarta;** plan 50 mm, höjd 50 mm (hårdgjorda ytor) annars 100 mm.
- **Grundkarta;** plan 50 mm, höjd 50 mm (hårdgjorda ytor) annars 100 mm.
- **Utstakning;** plan 20 mm, höjd 20 mm.
- **Lägeskontroll;** plan 50 mm, höjd 50 mm (hårdgjorda ytor) annars 100 mm.
- **Förrättningsmätning;** plan 20 mm.
- **Gränsutvisning;** plan 20 mm.

Dessa krav kan enkelt sorteras in i föreslagen intervallskala. Skalan harmonierar även väl med HMK-standardnivåerna i Tabell 2.6.

A.8 Lägesosäkerhet vs. antal värdesiffror

I Tabell A.8 anges lämplig upplösning i redovisningen av data i förhållande till deras standardosäkerhet. Med "upplösning" avses här antal *värdesiffror*, dvs. på vilken nivå man bör avrunda datavärdena.

Tabell A.8. Lämplig upplösning i förhållande till datas standardosäkerhet.

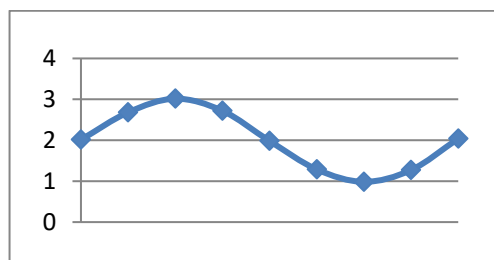
Standardosäkerhet	Lämplig upplösning
> 10 m	m
1 - 10 m	dm
0.1 - 1 m	cm
0.01 - 0.1 m	mm

Varför ska man då redovisa fler siffror än vad standardosäkerheten indikerar?

En gammal god, praktisk regel säger att man ska vara generös med siffrorna under beräkningsgången och inte avrunda förrän på slutet. Och geodata är ofta mätdata eller *grunddata* - en sorts "halvfabrikat" som ska användas, och kanske bearbetas, i nästa steg.

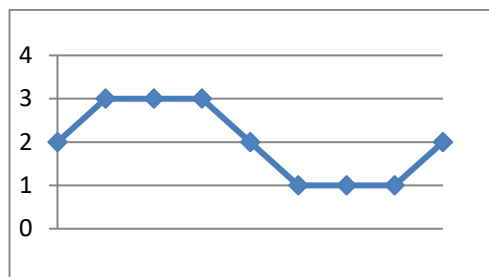
Exempel kan vara anpassning av ett lutande plan till ett laserpunktmoln i anslutning till en stöd- eller kontrollpunkt och interpolation för att ta fram profiler ur en höjdmmodell.

Höjddata t.ex. är oftast *korrelerade*, dvs. de ser ungefär ut som i Figur A.8.a. Variationerna är "långvågiga", dvs. överensstämmelsen **lokalt**, mellan närliggande datavärden, är god medan överensstämmelsen **regionalt** - det som standardosäkerheten mäter - är sämre.



Figur A.8.a. Interpolation mellan högupplösta datavärden.

I Figur A.8.a går det utmärkt att interpolera. Detta gäller dock endast om inte avrundningen har drivits för långt. Figur A.8.b innehåller samma värden som Figur A.8.a, men nu har varje datavärde avrundats till heltal, vilket tydligt försämrar resultatet.



Figur A.8.b. Interpolation mellan hårt avrundade utgångsvärden

Det teoretiska motivet är att om man avrundar för hårt och för tidigt så förstör man normalfördelningen. Den blir stympad och får andra egenskaper än de som vi normalt hänvisar till (signifikansnivåer, fel-fortplantningsformler etc.). Normalfördelningen kräver egentligen oändligt många decimaler.

Sammantaget bör man vara generös med antalet värdesiffror och rekommendationen i Tabell A.8 bör följas.

B Specifikationer, checklistor m.m.

I B.1 redovisas en innehållsförteckning för dataproduktspecifikationer (enligt [ISO 19131](#)) och en checklista för upprättandet av sådana. B.2 beskriver hur olika geodataaktörer bör läsa och tolka en dataproduktspecifikation. Dessutom ingår i denna bilaga en generell beskrivning av hela kontrollprocessen (B.3) och en översikt över olika kontrollmetoders användbarhet (B.4).

B.1 Att ta fram en dataproduktspecifikation

Rekommendationerna följer [SIS-TR 40:2012: Handbok för dataproduktspecifikation](#) ⁵⁾. Standarden beskriver delarna som ingående klasser i en databas medan handboken betraktar delarna som avsnitt i ett dokument. En dataproduktspecifikation (DPS) kan presenteras på olika sätt, som till exempel en pdf eller webbsida. En pdf bör följa ett gemensamt mönster med likadana rubriker, en webbsida bör ha klickbara rubriker till delarna enligt denna mall.

Innehåll

Innehållet i en DPS redovisas i Tabell B.1.a. De rubriker som inte är relevanta för dataprodukten ges värdet "ej relevant" alternativt "uppgift saknas" eller tas bort. Huvudrubrikerna och obligatoriska rubriker lämnas kvar.

Tabell B.1.a. *Dataproduktspecifikationens innehåll. Observera att siffrorna på huvudrubrikerna och huvudrubrikernas ordning inte är en del av standarden, det vill säga ordningen och siffrorna kan variera mellan olika DPS.*

Rubrik	Underrubrik	Beskrivning
1 Specifikationens omfattning		Anger dataproduktspecifikationens avgränsning.
2 Översikt ¹⁾		Sammanfattande obligatoriskt avsnitt för dataproduktspecifikationen. Rekommendation är att använda underrubrikerna.
2.1 Information om specifikationen ²⁾	Titel Datum Ansvarig part Språk Ämnesområde Syfte	Information om skapandet av dataproduktspecifikationen. Viss information kan stå i sidhuvud/sidfot.

⁵⁾ Handboken har nu uppdaterats till SIS-TR 40:2015, men eventuella effekter i bilagan av denna förändring har inte hunnit beaktas i HMK – Geodatakvalitet 2017

2.1.1 Termer och definitioner ¹⁾		Termer och deras definitioner, kan ersättas med en länk till källan.
2.1.2 Förkortningar ¹⁾		Förkortningar och dess fulla betydelser.
2.2 Sammanfattning av dataprodukten ¹⁾	Namn ¹⁾ Förkortning ¹⁾ Beskrivning av innehåll ¹⁾ Utsträckning Syfte Datakällor Produktionsprocesser Ajourhållning	Sammanfattning av dataprodukten eller informationsmängden. Viss information kan passa i ett dokumentets huvud. Förkortning anges om den existerar.
2.2.1 Kunders användning		Frivillig beskrivning; tänkt / faktisk användning och tänkta/faktiska kunder.
3 Omfattning ¹⁾		Beskrivning av datamängdens avgränsning, vilket kan vara i tid och rum, objekttyp, egenskapstyp, egenskapsvärde, rumslig representation med mera. En omfattning kan indelas i delomfattningar.
3.1 Hela datamängden ¹⁾	Identifiering ¹⁾ Hierarkisk nivå ¹⁾ Namn på hierarkisk nivå ¹⁾ Beskrivning av hierarkisk nivå ¹⁾ Utsträckning ¹⁾ Yttäcken ¹⁾ Ingående omfattningar	Beskrivning av omfattning av hela datamängden. Hierarkisk nivå enligt metadatastandard (ISO 19115-1), till exempel serie, datamängd och objekttyp. Yttäcken är till exempel raster, grid och TIN.
3.2 Delomfattning	Identifiering ¹⁾ Hierarkisk nivå ¹⁾ Namn på hierarkisk nivå ¹⁾ Förkortning Beskrivning av hierarkisk nivå ¹⁾ Utsträckning ¹⁾ Yttäcken ¹⁾ Ingående omfattningar Överordnad omfattning ¹⁾	Beskrivning av en del av datamängden. Kan göras i det avsnitt som berörs av omfattningen. Då kan identifiering utelämnas.

4 Identifiering ¹⁾	Titel ¹⁾ Alternativ titel Versionsnummer Sammanfattning ¹⁾ Syfte Ämnesområde ¹⁾ Metod för rumslig representation Rumslig upplösning Geografisk utsträckning ¹⁾ Kompletterande information	Grundläggande beskrivning av dataprodukten. Exempel på rumslig representation är: vektor, raster och grid. Exempel på ämnesområde är: navigering och hälsa. Rumslig upplösning kan vara av typen kartskalans nämnare eller upplösning på marken. Geografisk utsträckning kan vara omslutande rektangel, polygon eller textuell beskrivning.
5 Datainnehåll och struktur ¹⁾	Beskrivande text ¹⁾ Applikationsschema Objekttypskatalog ¹⁾ Omfattning ¹⁾	Utförlig beskrivning av innehållet och strukturen i dataprodukten, där applikationsschema (ISO 19109) och objekttypskatalog (ISO 19110) kan levereras som bilaga eller länk.
6 Referenssystem ¹⁾	Rumsligt referenssystem ¹⁾ Temporalt referenssystem Referenssystemets omfattning ¹⁾	Rumsligt referenssystem bör anges med EPSG-kod. Även indirekta referenssystem förekommer (se ISO 19112).
7 Kvalitetskrav ¹⁾		Beskrivning av kvalitetskrav på datamängden anges enligt ISO 19157 .
7.1 Datakvalitet	Fullständighet ¹⁾ Logisk konsistens ¹⁾ Lägesosäkerhet ¹⁾ Tematisk osäkerhet ¹⁾ Temporal kvalitet ¹⁾ Kvalitetsområde ¹⁾ (omfattning)	Beskriv det kvalitetsområde (geografiskt eller tematisk) som datakvalitetskraven gäller för. Ange de krav som ställs inom respektive kvalitetstema. För produkter kan ytterligare krav anges för prestanda, kapacitet och tillgänglighet.
8 Metadata ¹⁾		Beskriv de metadata som registreras.

9 Tillhandahållande ^{I)}	Sammanfattning Omfattning ^{I)}	Beskrivning hur dataprodukten tillhandahålls.
9.1 Leveransformat ^{I)}	Beteckning ^{II)} Version Specifikation Filstruktur Språk ^{II)} Teckenuppsättning	Dataformat för leverans.
9.2 Leveransmedium ^{I)}	Indelningsalternativ ^{II)} Medium Volym Övrig information	Medium för leverans enligt metadatastandard, till exempel dvd, ftp, och online.
9.3 Leveransrestriktioner		Frivilliga uppgifter om begränsningar och restriktioner av till exempel användbarhet, åtkomst och nyttjande.
9.4 Funktioner		Frivilliga uppgifter om funktioner hos en tjänst.
10 Datafångst	Sammanfattning Tillkomsthistorik	Frivilligt kapitel som kan innehålla lista på dataleverantörer.
11 Underhåll av data	Sammanfattning Underhållsfrekvens Omfattning	Frivilligt kapitel om underhåll. Underhållsfrekvens anges enligt ISO 19115-1 till exempel kontinuerlig, månatlig, årlig, periodisk med flera.
12 Presentationsregler	Hänvisning Omfattning	Hänvisning kan ske till presentationsregler på annan plats. Se ISO 19117 .
13 Övrig information		Kapitel för övrig information om dataprodukten.
14 Bilagor		Hänvisning till bilagor.
15 Referenser		Referenser till standarder och andra dokument.

^{I)} Obligatorisk uppgift enligt [ISO 19131](#)

^{II)} Obligatorisk uppgift enligt [SIS-TR 40:2012](#)

Checklista för framtagningen

I Tabell B.1.b visas en checklista för framtagning av en DPS. Listan utgör en utvidgning av beskrivningen i handboken [SIS-TR 40:2015](#), med utökade beskrivningar av tillvägagångssättet.

Tabell B.1.b. Checklista för framtagning av en dataproduktspecifikation.

Steg	Handling
1	Skapa ett dokument med: Namn, ansvarig organisation, specifikationens språk, datum och version kan skrivas i sidhuvud eller sidfot.
2	Formulera syftet med dataprodukten. Beskriv tänkt nytta och vilka behov dataprodukten uppfyller.
3	Dokumentera dataproduktens innehåll och struktur med applikationsschema (UML rekommenderas) och objekttypskatalog.
4	Beskriv dataproduktens tänkta kunder och användning i kapitlet <i>Översikt</i> .
5	Beskriv datainsamling och processer för uppbyggnad.
6	Beskriv förvaltning och ajourhållning av dataprodukten.
7	Beskriv presentationsregler om det är relevant.
8	Beskriv hur datamängden görs tillgänglig, med eventuella restriktioner.
9	Ange geografiska referenssystem helst med kod (EPSG) och klartext.
10	Beskriv grundläggande metadata samt de metadata som tillhandahålls vid leverans.
11	Ange <i>Titel</i> med dataproduktens namn eller beteckning, ange <i>ämnesområde</i>
12	Beskriv delomfattningar med avvikande kvalitetskrav om de förekommer.
13	Sätt datakvalitetskrav för att säkerställa att syftet med dataprodukten kan uppnås.
14	Sammanfatta allt i översikten.

B.2 Att läsa en dataproduktspecifikation

Det finns flera olika roller i geodataverksamheten. Alla dessa har skilda behov av information och kommer därför att förhålla sig till en dataproduktspecifikation (DPS) på olika sätt. De viktigaste rollerna bedöms vara beställaren, utföraren/producenten, IT-utvecklaren och användaren.

Beställaren

Beställaren är ägare av dataprodukten och den som skriver och förvaltar dataproduktspecifikationen – i egen regi eller i samverkan med andra beställare av samma produkt. För detta måste ett antal beslut fattas, till exempel beträffande datainnehåll, kvalitetskrav och leveransutformning.

Beställaren följer arbetsgången i Tabell B.1.b för att ta fram en ny dataproduktspecifikation eller nyttjar befintlig dataproduktspecifikation som en del av en teknisk specifikation vid upphandling eller internbeställning av produktion samt vid kontroll av en levererad dataprodukt.

Utföraren

Utföraren genomför produktionen på specificerat sätt, med specificerad kvalitet, och levererar enligt specificerad leveransutformning.

Utföraren läser om dataproduktens innehåll, med dess kvalitetskrav, samt vilka metadata som ska ingå i leveransen. I förvaltningsskedet är kapitel om underhåll särskilt intressanta.

IT-utvecklaren

IT-utvecklaren tar fram program och datorsystem för insamling, lagring, bearbetning, analys, presentation, tillhandahållande med mera.

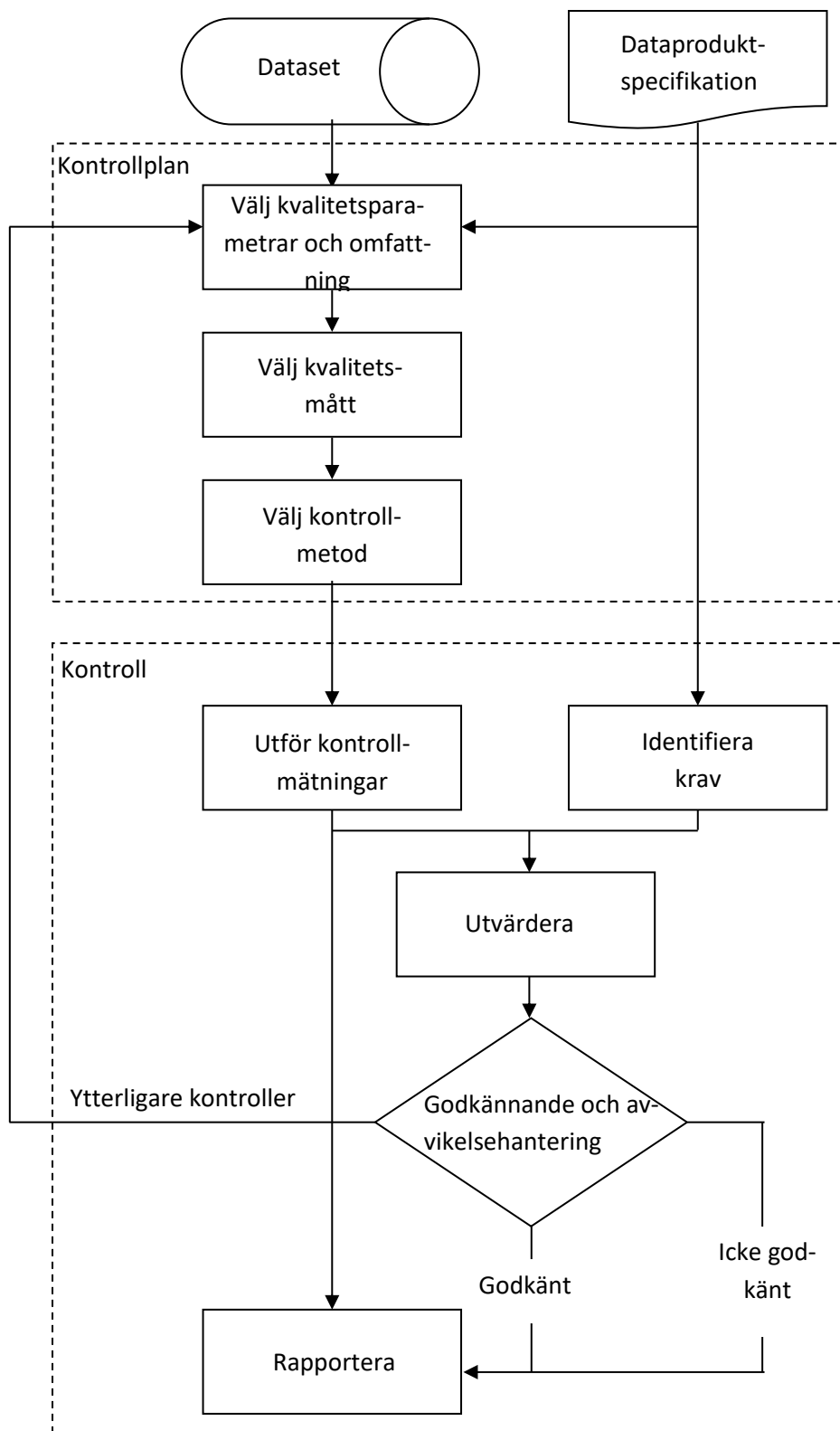
Läser applikationsscheman och information om tillhandahållande samt metadata.

Användaren

Användaren ska kunna förstå datamängden för att bedöma dess användbarhet för sina egna direkta behov eller för vidareförädling.

Söker efter rätt informationsmängd genom metadata och de inledande kapitlen. Även tillhandahållande och datakvalitetskrav tas i beaktande för att avgöra dataproduktens lämplighet för ändamålet.

B.3 Kontrollprocessen



Figur B.3. Schematisk översikt över kontrollprocessen. (Direkt översatt från den norska standarden Geodatakvalitet)

En översikt över kontrollprocessen redovisas i Figur B.3. Processen består av momenten "Upprätta kontrollplan" och "Genomför kontroll", se Tabell B.3.

Både figuren och tabellen utgör (fria) översättningar från den nya norska standarden "Geodatakvalitet".

Tabell B.3. Kontrollprocessens steg. Fri, kortfattad översättning från den nya norska standarden "Geodatakvalitet".

Upprätta kontrollplan	
Välj kvalitetsparametrar och omfattning	Till exempel <i>lägesosäkerhet</i> och <i>fullständighet</i> , se Tabell 2.7; kontrollens omfattning måste stå i rimlig proportion till datamängdens storlek.
Välj kvalitetsmått	Till exempel <i>standardosäkerhet</i> som mått på lägesosäkerheten eller <i>brist/övertalighet</i> , <i>andel fel</i> , som mått på fullständigheten.
Välj kontrollmetod	Till exempel mätning med nätverks-RTK för att kontrollera lägesosäkerhet och fältkontroll av fullständigheten.
Genomför kontroll	
Utför kontrollmätningar	Bestäm värden på de aktuella kvalitetsmåten.
Identifiera krav	Vanligen i form av toleranser.
Utvärdera	Jämför värdena på kvalitetsmåten från kontrollmätningen med toleranserna.
Godkännande och avvikelshantering	Ställningstagande om godkännande eller underkännande, hantering av avvikelser samt beslut om eventuell utvidgad kontroll.
Rapportera	Redovisa, med tonvikt på avvikelser.

B.4 Olika kontrollmetoders användbarhet

Kontrollmetoderna fungerar olika bra i olika sammanhang, se Tabell B.4.

Tabell B.4. Översikt över ett antal metoders användbarhet för kontroll av olika kvalitetsparametrar. Översättning från, och viss anpassning av, en tabell i den norska standarden "Geodatakvalitet".

Kvalitets-tema	Kvalitetsparametrar	Kontrollmetoder								
		Kontroll av dokumentation	Automatiserade kontroller	3D-Visualisering	Visuell kontroll mot ortofoto	Kontroll mot andra data	Visuell kontroll	Kontrollmätning	Fältkontroll	Kontrollmätning
Fullständighet	Brist	+			+	++	++		++	
	Övertalighet	+			+	++	++		++	
Logisk konsistens	Konceptuellt	+	++							
	Domän	+	++							
	Format	+	++							
	Topologi	+	++	+		+				
Lägesosäkerhet	Absolut	+			+	+	+	++		++
	Lokal	+		+				++		++
	Rasterdata	+			+	+		++		++
	(Tillförlitlighet)	+						++		++
Tematisk osäkerhet	Klassificering	+			+	++	+		++	
	Kvalitativa attribut	+				++			++	
	Kvantitativa attribut	+				++		+	++	++
Temporal kvalitet	Tid	+	++	+						
	Temporal konsistens	+	++							
	Temporal validitet	+				++				
Användbarhet	Lämplighet	+								
	Aggregering	+	++							

++ Metoden är användbar
+ Metoden kan vara användbar

C Ett tillämpningsexempel

I denna bilaga redovisas ett större tillämpningsexempel, som anknyter till en ganska ambitiös kontrollmetod.

C.1 Dataproduktspecifikationens krav

En tänkt dataproduktspecifikation anger följande specificerade datakvalitet:

- **Tematisk osäkerhet**, byggnader: $p_o = 0,03$ (3 % felaktigt klassificerade objekt)
 - **Fullständighet**, stödmurar: $p_o = 0,05$ (5 %)
 - **Lägesosäkerhet** (i plan), gränspunkter: $\sigma = 20$ mm (standardosäkerhet)
- Ingen **systematisk avvikelse**: $\mu = 0$
Grova fel: $p_o = 0,01$ (1 %)

Tre kontrollområden, som tillsammans täcker datamängdens geografiska utbredning, läggs ut. En objekttyp testas i varje kontrollområde på det sätt som redovisas i Tabell C.1. Stickprovsstorlekarna har beräknats med hjälp av Tabell 3.4.

Tabell C.1. Stickprovsstorlekar för tillämpningsexemplet. Vad gäller kontrollen av gränspunkter bör dock stickprovet innehålla 20 objekt eftersom kontrollen av lägesosäkerhet, systematik och grova fel är samordnad.

Objekttyp	Typ av kontroll	Antal förekomster i kontrollområdet	Stickprovsstorlek (n)
Byggnader	Tematisk osäkerhet, objekt-klassificering (kvalitativ)	856	80
Stödmurar	Fullständighet (kvalitativ)	58	13
Gränspunkter	Lägesosäkerhet och systematik (kvantitativ), grova fel (kvalitativ)	132	10 (20) 20

C.2 Test av tematisk osäkerhet

För ett stickprov på 80 objekt och med den specificerade kvaliteten $p_o = 0,03$ (3 %) blir toleransen enligt Tabell A.6

Tolerans, tematisk osäkerhet: antalet felklassificeringar < 6 st.

Vid kontrollen fann man 5 "Uthus" som hade klassificerats som "Bostadshus". Eftersom $5 < 6$ så godkänns det aktuella kontrollområdet i detta avseende. Felklassificeringarna ska dock rättas till.

C.3 Test av fullständighet

För ett stickprov på 13 objekt och med den specificerade kvaliteten $p_o = 0,05$ (5 %) blir toleransen enligt Tabell A.6

Tolerans, fullständighet: antalet brister < 3 st.

Vid kontrollen saknades 3 stödmurar som borde ha varit med enligt dataproductspecifikationen. Eftersom $3 \geq 3$ så underkänns hela det aktuella kontrollområdet i detta avseende. Komplettering med de saknade stödmurarna ska också ske.

C.4 Test av grova fel, systematik och lägesosäkerhet

Testen av lägesosäkerhet, systematik och grova fel samordnas. Samtliga utgår från dataproductspecifikationens krav på standardosäkerheten i plan ⁶ ($\sigma = 20\text{mm}$). Detektering av eventuella grova fel bör ske först.

Grova fel

Avvikelse större än 3σ definieras som grova fel. Det ger i vårt fall gränsvärdet 60 mm för radiella fel. För ett stickprov på 20 objekt och med den specificerade kvaliteten $p_o = 0,01$ (1 %) blir toleransen enligt Tabell A.6

Tolerans, grova fel: antalet avvikelser större än $3\sigma < 2$ st.

Vid kontrollen hittades ett grovt fel. Eftersom $1 < 2$ så godkänns hela det aktuella kontrollområdet i detta avseende. Det grova felet ska dock rättas till, eller åtminstone rapporteras.

Systematik

Systematiken kontrolleras genom att studera den *absoluta* (1D) eller *radiella* (2D och 3D) medelavvikelsen. Denna betecknas $\bar{\Delta}$ och beräknas enligt (se fotnot i Bilaga A.2)

$$- \bar{\Delta} = |\bar{\varepsilon}| \quad (\text{medelskift}) \quad (1D)$$

$$- \bar{\Delta}_R = \sqrt{\bar{\varepsilon}_N^2 + \bar{\varepsilon}_E^2} \quad (2D)$$

$$- \bar{\Delta}_R = \sqrt{\bar{\varepsilon}_N^2 + \bar{\varepsilon}_E^2 + \bar{\varepsilon}_H^2} \quad (3D)$$

⁶ Om kontrollmetoden inte kan betraktas som felfri så ska dess standardosäkerhet $\sigma_{kontroll}^2$ reduceras bort via formeln $s_{korr} = \sqrt{s^2 - \sigma_{kontroll}^2}$

För ett stickprov på 20 objekt, med specifikationen att ingen systematik ska finnas ($\mu=0$), blir gränsvärdet $\bar{\Delta}_R \leq s_{plan} \cdot t / \sqrt{n}$. För $s_{plan} = 21,1$, $n = 20$ och $t / \sqrt{20} = 1,80 / \sqrt{20} = 0,40$ (Tabell A.5.c) får vi toleransen

Tolerans, systematik: radiell medelavvikelse

$$\bar{\Delta}_R \leq 21,1 \cdot 0,40 = 8,4 \text{ mm.}$$

Vid kontrollen skattades medelavvikelsen till $\bar{\Delta}_R = 25,0$ mm. Eftersom $25,0 > 8,4$ så underkänns hela det aktuella kontrollområdet i detta avseende. Orsaken till den systematiska avvikelsen i plan måste utredas.

Lägesosäkerhet

För ett stickprov på 20 objekt, med den specificerade kvalitén att standardosäkerheten i plan (2D) ska vara max 20 mm, blir toleransen enligt Tabell A.5.b

Tolerans, standardosäkerhet:

$$s_{plan} \leq \sigma \cdot \sqrt{F_{0,05,2,(20-1),\infty}} \approx 20 \cdot (0,96 + (38)^{-0,4}) = 20 \cdot 1,19 = 23,8 \text{ mm}$$

där s_{plan} är stickprovets beräknade standardosäkerhet. Ev. grova fel från föregående steg ska först tas bort ur materialet. Vid kontrollen skattades standardosäkerheten i plan till $s_{plan} = 21,1$ mm. Eftersom $21,1 \leq 23,8$ så godkänns det aktuella kontrollområdet i detta avseende.

D Norskt register över kvalitetsmått

Denna bilaga är ett utdrag ur den nya norska standarden *Geodatakvalitet*, som har översatts till svenska. Den beskriver ett norskt register över kvalitetsmått för att "mäta" datakvalitet. Framställningen följer standarden [ISO 19157](#) mycket väl och innehåller flera konkreta förslag på kvalitetsmått inom samtliga kvalitetsteman och för samtliga kvalitetsparametrar. I ISO-standarderna finns många illustrativa figurer som ett värdefullt komplement för att fullt ut förstå de olika måtten.

Bilagan kompletterar (och överlappar) avsnitt 2.7, och ger en samlad beskrivning av kvalitetsteman, -parametrar och -mått.

D.1 Inledning

Enligt datakvalitetsmodellen i [ISO 19157](#) kan resultaten anges på tre olika sätt:

- *kvantitativa resultat* (quantitative result); ett (eller flera) beräknade värden som anger datakvaliteten
- *grad av överensstämmelse* (conformance result); uppgift om i vad mån datakvalitetskraven i den aktuella dataproduktspecifikationen är uppfyllda
- *beskrivande resultat* (descriptive result); används där det är önskvärt att säga något om datakvaliteten, utan att det är möjligt att kvantifiera resultatet. I samtliga fall där kvantifiering kan ske bör därför alternativ 1 (kvantitativa resultat) användas.

D.2 Fullständighet

Fullständighet betyder "faktiska förekomster i en datamängd jämfört med dem som borde ha varit med", d.v.s brist eller övertalighet.

Kommentar: Används främst i jämförelse med "verkligheten". I jämförelse med kraven i en informationsmodell används företrädesvis *logisk konsistens*.

Exempel på kvalitetsmått:

- andelen (eller antalet) saknade objekt, brist
- andelen (eller antalet) övertaliga objekt
- antalet duplicerade objekt (antal dubbla instanser).

D.3 Logisk konsistens

Med *logisk konsistens* avses förhållandet mellan de logiska regler som gäller för en datamängd och datamängden i sig.

Konceptuell konsistens

Konceptuell konsistens syftar på hur väl reglerna för det konceptuella (begreppsmässiga) schemat följs.

OBS: Denna kvalitetsparameter kan ses som en summering av flera olika krav, till exempel domänkrav (se domänkonsistens).

Exempel på kvalitetsmått:

- antalet enheter i datauppsättningen som inte är i enlighet med de regler som anges i det tillhörande konceptuella schemat
- antalet felaktiga överlappningar mellan ytor i datamängden
- antalet objekt som ska ha konstant höjdvärde i samtliga geometripunkter, men som bryter mot detta krav
- antalet objekt som ska ha monotont avtagande höjdvärden, men som bryter mot detta krav; höjdvärdena för punkterna utefter en bäck bör till exempel vara monotont avtagande i vattnets fallriktning.

Domänkonsistens

Med *domänkonsistens* avses i vad mån olika typer av värden håller sig inom sina respektive *värdeområden* (tillåtna värden).

OBS: Det behöver inte röra sig om numeriska storheter. I en informationsmodell är det t.ex. angett vilka namn som får användas, och då ingår det i analysen av domänkonsistens att kontrollera att inga andra namn används.

Exempel på kvalitetsmått:

- andelen enheter som matchar domänen
- andelen enheter som inte matchar domänen.

Formatkonsistens

Formatkonsistens syftar på relationen mellan en datamängd och dess fastställda lagringsstruktur.

Exempel på kvalitetsmått:

- andelen fysiska strukturkonflikter.

Topologisk konsistens

Topologisk konsistens anger hur korrekt datamängdens topologi är i förhållande till dataproduktspecifikationens topologikrav. Till exempel ska en bäck som rinner ut i en sjö sluta vid sjöns definierade strandlinje, och referenspunkten för en polygon (yta) ska ligga innanför begränsningslinjen.

Exempel på kvalitetsmått:

- antalet saknade konnektioner på grund av för korta linjer (Eng. *undershoots*)
- antalet felaktiga konnektioner på grund av för långa linjer (Eng. *overshoots*)
- antalet felaktiga småpolygoner (Eng. *slivers*)
- antalet felaktiga egenkorsningar (Eng. *loops* eller *invalid selfintersects*)
- antalet felaktiga självöverlappningar, till exempel: "fram- och tillbakalinjer" och dubbelpunkter (Eng. *kickbacks* eller *invalid selfintersects*)
- antalet felaktiga enkelnoder (Eng. *incorrect 1-nodes*)
- antalet felaktiga länkkorsningar, det vill säga att länkar korsar varandra utan att en fysisk skärningspunkt läggs ut (Eng. *crossing edges*)
- procentandelen fel i förhållande till komplett yttäckning (av slutna polygoner), det vill säga arealen "hål" i procent av den totala areal som skulle ha varit täckt (Eng. *(in)complete coverage*).

D.4 Lägesosäkerhet

Med *lägesosäkerhet* avses osäkerhet i en positionsangivelse, dvs. hur bra ett objekts rumsliga läge är fastställt i förhållande till verkligheten (facit).

Kvalitetsparametrarna är *absolut lägesosäkerhet* och *lokal lägesosäkerhet* (se Tabell 1) samt *lägesosäkerhet i rasterdata*.

Exempel på kvalitetsmått (för samtliga kvalitetsparametrar):

- procentandelen grova fel (Eng. *gross errors*)
- systematisk avvikelse i höjd (Eng. *bias, 1D*)
- systematisk avvikelse i 2D/3D (Eng. *bias, 2D/3D*)
- standardavvikelse (Eng. *standard deviation*).

D.5 Tematisk osäkerhet

Med *tematisk osäkerhet* menas osäkerheten i kvantitativa attribut och korrektheten i icke-kvantitativa attribut – samt i objektens klassificering och relationer.

Klassificeringsosäkerhet

Med *klassificeringsosäkerhet* avses riktigheten i objektklassificeringen eller i objektens kvalitativa egenskaper.

Exempel på kvalitetsmått:

- andelen felklassificerade objekt (Eng. *misclassification rate*)
- relativ felklassificeringsmatris/förväxlingsmatris (Eng. *relative misclassification matrix*)
- kappakoefficient.

Tematisk osäkerhet, kvalitativa attribut

Den *tematiska osäkerheten för kvalitativa attribut* avser riktigheten i icke-kvantitativa attributvärden.

OBS: kvalitativa attribut skiljer sig på ett sätt där rangordning inte är möjlig (nominalskala).

Exempel på kvalitetsmått:

- andelen felaktiga attributvärden (Eng. *rate of incorrect attribute values*).

Tematisk osäkerhet, kvantitativa attribut

Den *tematiska osäkerheten för kvantitativa attribut* avser osäkerheten i mätbara attributvärden.

OBS: kvantitativa attribut kan mätas på en skala (ordinalskala, intervallskala eller kvotskala).

Exempel på kvalitetsmått:

- standardavvikelse (Eng. *standard deviation*)
- systematisk avvikelse (Eng. *bias*)
- andelen (eller antalet) grova fel (Eng. *gross errors*).

D.6 Temporal kvalitet

Temporal kvalitet är datakvalitetssegenskaper som definierar tid eller tidsberoenden mellan objekt.

Tidsosäkerhet

Med *tidsosäkerhet* avses osäkerheten i tidsmätningen.

Exempel på kvalitetsmått:

- standardavvikelse (Eng. *standard deviation*)
- systematisk avvikelse (Eng. *bias*)
- antalet (eller andelen) grova fel (Eng. *gross errors*).

Temporal konsistens

Med *temporal konsistens* avses hur korrekt tidsordningen – kronologin – är för ordnade händelser eller sekvenser.

Exempel på kvalitetsmått:

- antalet kronologiska fel.

Temporal validitet

Med *temporal validitet* menas tidsuppgifters giltighet.

Exempel på kvalitetsmått:

- antalet uppgifter utanför sin värdeomän (Eng. *items not in conformance with their value domain*).

D.7 Användbarhet

Med *användbarhet* avses en datamängds lämplighet för en viss tillämpning och/eller dess överensstämmelse med dataproduktspecifikationen.

Exempel på kvalitetsmått:

- antalet (andelen) krav i dataproduktspecifikationen som datamängden inte uppfyller (Eng. *data product specification fail count/rate*)
- antalet (andelen) krav i dataproduktspecifikationen som datamängden uppfyller (Eng. *data product specification pass count/rate*).

D.8 Generellt kvalitetsmått

Kvalitetsmättet nedan kan användas på alla kvalitetsparametrar. Av praktiska skäl har det inte kopierats in under varje parameter utan anges bara här.

- uttalande om datakvaliteten i form av text (Eng. *data quality statement*).

OBS. Uttalandet bör om möjligt delas upp och sorteras in under de olika kvalitetsparametrarna.

Exempel på en utsaga om attributosäkerheten:

- Datamängden innehåller objekt som har genomgått en hög grad av tematisk generalisering (sammanslagning/ gruppering).

E Statistik, introduktion

E.1 De viktigaste statistiska storheterna

De centrala storheterna inom beskrivande statistik är medelvärde och standardosäkerhet.

Medelvärdet brukar betecknas \bar{x} och beräknas som

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

ur n stycken värden x_i .

En skattning av *standardosäkerheten* (σ) ges sedan av uttrycket

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Inom geodatavärdet är även storheten

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

vanligt förekommande. Sambanden mellan dessa statistiska storheter beskrivs i Bilaga A till [HMK-TR 2015:1](#), se även avsnitt 3.3 i detta dokument.

E.2 Terminologi

Den terminologiska grunden för kontroll av geodatakvalitet redovisas i Tabell E.2.

Tabell E.2. Grundläggande terminologi för kontroll av datakvalitet.

Krav	Specificerat gränsvärde i en dataproduktspecifikation, se avsnitt 4.4. Datamängdens uppmätta kvalitet ska inte, signifikant, vara sämre än detta krav.
Uppmätt kvalitet	Värdet på det <i>kvalitetsmått</i> som har uppmätts/beräknats för den del av datamängden som har kontrollerats.
Tolerans	Kravet justerat för <i>statistisk osäkerhet</i> – till exempel osäkerheten vid stickprovstagning och i kontrollmetoden.
Kvalitetskontroll	Arbetet med att undersöka om en datamängd har den kvalitet som specificerats. Det vill säga <i>uppfylls kraven</i> i dataproduktspecifikationen? Det sker genom jämförelse mellan uppmätt datakvalitet och tolerans.

E.3 Kontrollens trovärdighet

I enstaka fall kan kontrollen av datakvalitet omfatta alla förekomster av en viss objekttyp i ett kontrollområde. Då jämförs den uppmätta kvaliteten direkt med kravet i dataproduktspecifikationen. Det vanligaste är dock att kvalitetsmåttet bestäms genom *stickprovskontroll*.

Resultatet jämförs då i stället med aktuell *tolerans*. I bedömningen av om en datamängd håller avsedd kvalitet måste man även ta hänsyn till stickprovsurvalets och kontrollmetodens osäkerhet. Det betyder att, även om kontrollresultatet är sämre än kravet, så kan datamängden i sig vara av tillräckligt hög kvalitet.

Kvalitetsmålet måste nämligen vara *signifikant* sämre än kravet för att datamängden ska underkännas. Denna signifikans mäts via toleransen. (Kvalitetskontroll hanteras vidare i kapitel 3.)

E.4 Kravet på signifikans

Signifikans är ett siffermässigt uttryck för trovärdigheten i kontrollen. Signifikansen uttrycks vanligen i % och anger sannolikheten för att slutsatsen av kontrollen är felaktig. HMK tillämpar en *signifikansnivå* på 95 %, vilket motsvarar en *risknivå* på 5 %.

Kravet kan ses från producentens eller användarens perspektiv:

- **Producenten** kan kräva att det är ≤ 5 % risk att en leverans blir **underkänd**, trots att den håller tillräcklig datakvalitet (*Producers risk* eller *Typ I-fel*).
- **Användaren** (kunden) kan kräva att det är ≤ 5 % risk att en leverans som skulle ha varit underkänd blir **godkänd** (*Consumers risk* eller *Typ II-fel*).

HMK utgår från producenten, det vill säga att risken för Typ I-fel är ≤ 5 %. Det betyder alltså att det ska finnas ett statistiskt underlag för att påstå att produkten är sämre än kravet.

Antalet objekt som kontrolleras påverkar kontrollens trovärdighet. Det är svårare att få en datamängd förkastad om stickprovet innehåller få objekt, eftersom toleransen då blir mer generös.

E.5 Konfidensintervall

Kvalitetsmålet skattas ur stickprovet, medan det "sanna" värdet, som avser samtliga förekomster av den aktuella objekttypen, kan vara större eller mindre.

Ett *konfidensintervall* är ett intervall som innehåller det "sanna" värdet med en viss sannolikhet – till exempel 95 % signifikansnivå, somt är det vanligaste värdet och tillämpas normalt i HMK.

E.6 Hypotesprövning

För att testa om kvalitetsmättet är signifikant sämre än toleransen utförs ett statistiskt *hypotestest*.

Det baseras på två hypoteser, som prövas i tur och ordning:

- Nollhypotesen H_0 : Kraven uppfylls
- Alternativhypotesen H_1 : Kraven uppfylls ej

Man börjar med att testa hypotesen H_0 . Endast om den förkastas så accepteras i stället alternativhypotesen H_1 , det vill säga underkänt.

E.7 Kvalitetsredovisning vs. kvalitetskontroll

Konstruktion av konfidensintervall och hypotesprövning är närbesläktade operationer. Som beskrivits ovan gäller dock normalt att

- konfidensintervall används vid redovisning av faktisk kvalitet
- hypotesprövning används vid kontroll av att kvaliteten är den specificerade.

Se vidare exemplen E.7.a-b vad gäller relationen mellan konfidensintervall och hypotesprövning – eller mellan redovisning och kontroll.

Exempel E.7.a: Ur det skattade medelvärdet \bar{x} och den kända standardosäkerheten σ kan det två-sidiga konfidensintervallet

$$\left[\bar{x} - 1,96 \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + 1,96 \cdot \frac{\sigma}{\sqrt{n}} \right]$$

konstrueras. Det innehåller det sanna medelvärdet (väntevärdet) μ med sannolikheten 95% vid normalfördelning.

Exempel E.7.b: Vid till exempel test av systematik vill man ibland veta om medelvärdet/väntevärdet $\mu \neq 0$. Det kan man göra genom att kontrollera om nollan (0) ligger inom intervallet i det förra exemplet. Så är fallet om absolutbeloppet:

$$|\bar{x}| \leq 1,96 \frac{\sigma}{\sqrt{n}} \text{ (tolerans)}$$

som alltså blir vårt hypotestest av hypotesen $H_0: \mu = 0$. Om $|\bar{x}|$ är större än denna tolerans så förkastas H_0 och vi kan konstatera att en signifikant (95%) systematik finns, det vill säga $H_1: \mu \neq 0$ gäller.

För att åstadkomma enhetlighet och enkelhet används **hypotesprövning** genomgående i HMK, eftersom det vanligen rör sig om jämförelse med en tolerans.