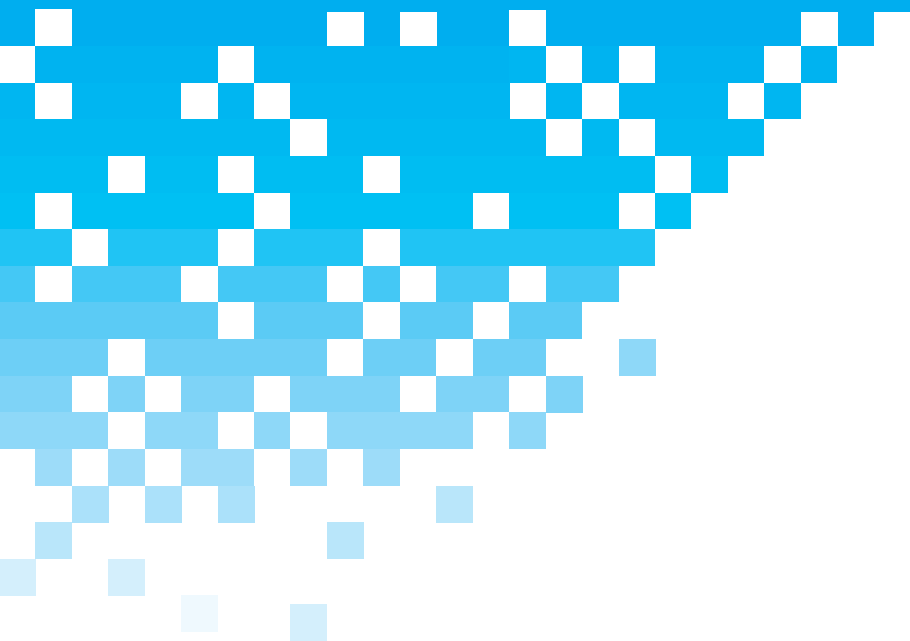


HMK
- handbok i mät- och kartfrågor

Geodatakvalitet

2014



Förord 2014

Detta är ett nytt HMK-dokument, som tagits fram under 2014. Här har det mesta som handlar om geodatakvalitet samlats, och till del brutits ut ur HMK-Introduktion 2013.

Det innebär att de inledande HMK-delarna nu består av två separata dokument

- HMK-Introduktion 2014, renodlad till en mer regelrätt introduktion
- HMK-Geodatakvalitet 2014, en kraftigt utvidgad beskrivning av datakvalitetsfrågorna inom området.

Dessa båda dokument kompletterar varandra, och HMK-Geodatakvalitet utgör fortfarande en del av introduktionen in i HMK. Det är alltså inte en handbok i egentlig mening – snarare ett referensdokument, som alla kan slå i när man undrar över något beträffande datakvalitet och andra kvalitetsfrågor.

Arbetet med HMK-Geodatakvalitet har huvudsakligen utförts av undertecknad, Thomas Lithén, Gunhild Lönnberg och Torsten Svärd, Lantmäteriet. Två externa granskningar har utförts under innevarande sommar respektive höst.

Östersund 2014-12-17

Clas-Göran Persson
Dokumentansvarig

[Samlade förord](#)

Innehållsförteckning

1	Inledning	5
1.1	Detta dokument.....	5
1.2	HMK:s roll.....	5
1.3	Inhemska influenser.....	6
1.4	Internationella standarder och bidrag.....	7
1.5	Det norska standardiseringsarbetet.....	8
2	Statistik, introduktion	10
2.1	De viktigaste statistiska storheterna.....	10
2.2	Terminologi.....	10
2.3	Kontrollens trovärdighet.....	11
2.4	Kravet på signifikans.....	11
2.5	Konfidensintervall.....	11
2.6	Hypotesprövning.....	12
2.7	Kvalitetsredovisning vs. kvalitetskontroll.....	12
2.8	Mätosäkerhet - GUM.....	13
3	Datakvalitetsfilosofi	15
3.1	Förändrad syn på kvalitetskontroll.....	16
3.2	HMK-standardnivåer.....	18
4	Datakvalitet	21
4.1	Definition av geodatakvalitet.....	21
4.2	Datakvalitetsmodell och informationsmodellering.....	22
4.3	Metadata.....	23
4.4	Dataproduktspecifikationen.....	23
4.5	Kvalitetsteman och kvalitetsparametrar.....	24
4.6	Relationer mellan olika kvalitetsparametrar.....	27
4.7	Kvalitetsparametrar vs. kvalitetsmått.....	29
4.8	Intervallskala för lägesosäkerhet.....	30
4.9	Begreppet kontrollerbarhet.....	31
5	Kontroll av geodata	33
5.1	Olika typer av kontroller.....	33
5.2	HMK:s principer för toleranser (felgränser).....	34
5.3	Systematisk avvikelser, standardavvikelse, RMS och grova fel.....	37
5.4	Stickprov och kontrollområden.....	39
6	Läs mer	41

A	Exempel på kontrollmätningar och toleranser ...	42
A.1	Förenklad kvalitetskontroll	42
A.2	Kontroll av mätosäkerheten i Nätverks-RTK.....	43
A.3	Kontroll av utsättning	45
A.4	Kontroll av geoidmodell.....	47
A.5	Test av grova fel, mätosäkerhet och systematik i 1 D, 2D och 3D	48
A.6	Kontroll av grova fel och fullständighet/objekt- klassificering.....	51
B	Specifikationer, checklistor m.m.	53
B.1	Att ta fram en dataproduktspecifikation	53
B.2	Att läsa en dataproduktspecifikation.....	58
B.3	Kontrollprocessen	59
B.4	Olika kontrollmetoders användbarhet	61
C	Ett större tillämpningsexempel	62
C.1	Dataproduktspecifikationens krav	62
C.2	Test av objektklassificering	62
C.3	Test av fullständighet	63
C.4	Test av grova fel, lägesosäkerhet och systematik.....	63
D	Norskt register över kvalitetsmått	65
D.1	Inledning	65
D.2	Fullständighet	65
D.3	Logisk konsistens	65
D.4	Lägesnoggrannhet.....	67
D.5	Tematisk noggrannhet	67
D.6	Temporal noggrannhet	68
D.7	Användbarhet	69
D.8	Generellt kvalitetsmått.....	69

1 Inledning

1.1 Detta dokument

I dokumentet ges en ganska ingående beskrivning av begreppet *geodatakvalitet*, till exempel terminologin och dess teoretiska grundval. Framställningen utgår från det traditionella HMK-upplägget men är anpassat till nationella och internationella standarder.

I tillägg tillhandahåller vi ett "smörgåsbord" bestående av ett antal metoder för kontroll av geodatakvalitet – från de enkla, till de mer strikta, "statistiskt korrekta". Dessa har samlats i en bilaga – för att öka överblicken och underlätta metodvalet utifrån vars och ens behov, ambitionsnivå med mera.

Därigenom blir dokumentet något mitt emellan lärobok och handbok, till vilket man refererar från övriga HMK-dokument. Genom att samla ihop datakvalitetsfrågorna hoppas vi kunna bidra till enhetlighet och samsyn mellan geodataområdets olika aktörer: beställaren, utföraren, IT-utvecklaren, användaren med flera (se Bilaga B.2).

1.2 HMK:s roll

Vi ser HMK:s roll som att utgöra en brygga mellan den formella standardiseringsverksamheten och det vardagliga arbetet för beställare, utförare och användare. Det innebär förtydliganden och konkretiseringar såväl som ställningstaganden i vägvalsfrågor och, i viss mån, anpassning till nationella förhållanden.

Standarder är nämligen inte så samordnade och kompletta som man kanske tror:

- De överlappar varandra, och det finns till och med motsägelser mellan dem.
- Det finns delar som inte täcks, eller delar som stannar vid det teoretiska och inte ger tillräckligt tydliga praktiska anvisningar ¹
- Det finns behov av nationella, terminologiska anpassningar.

Dessutom finns det andra nationella, "kulturella" avvikelser från den kompromiss som internationell standardisering ju alltid innebär, till exempel informella standarder, branschstandarder och de facto-standarder – som har stor spridning och har använts länge.

¹ Man kan gå så långt som att säga att ISO medvetet stannar på en abstrakt nivå, som måste konkretiseras. Andra standardiseringsorgan – till exempel OGC – tar fram färdiga tillämpningsstandarder.

De rutiner som de stora beställarna – exempelvis Trafikverket i Sverige – tillämpar påverkar också den nationella praxisen inom området.

Ser man historiskt på det hela så publicerades de tidiga, svenska standardiseringsansatserna från slutet av 1980-talet först i HMK-serien från 1993-95. Och hanteringen av geodatakvalitet i HMK-Databaser från 1994 skiljer sig inte nämnvärt från de formella standarder som kom senare. Till exempel fanns det centrala konceptet med *dataproduktspecifikation* med redan där, även om det då benämndes *databasspecifikation* och främst avsåg den geodatabas ur vilken produkter framställdes.

Vad gäller just terminologifrågorna så kan det vara på sin plats att även motivera existensen av den egna [HMK – Ordlista och förkortningar](#) (december 2014). Där finns förklaringar till fackuttryck och till de förkortningar som används. Syftet med ordlistan är göra HMK-dokumenterna begripliga genom att ge en samlad och enhetlig beskrivning av använda termer, begrepp och förkortningar.

Inom HMK-projektet finns inga ambitioner att vara generellt normerande men vi försöker naturligtvis använda vedertagen terminologi inom berörda områden.

Att även de vanligaste förkortningarna finns med är en ren service; läsaren ska inte behöva leta dessa i andra dokument.

Vi menar sammantaget att HMK behövs – parallellt och som komplement. Detta styrks bland annat av att man inom geodataområdet i Norge har jobbat på precis det sättet med stor framgång, se slutet av detta avsnitt. HMK kan också utgöra inkörsport till de formella standarderna och vi ser ingen som helst motsättning mellan de två; båda parter har en effektiv geodataverksamhet som yttersta mål.

1.3 Inhemska influenser

Detta dokument är en utvidgning av

- Lantmäteriet, 2013. [HMK-Introduktion](#) 2013.

vad gäller geodatakvalitet. Tidiga influenser är

- Lantmäteriverket, 1996. Handbok till mätningsskugörelsen: [HMK-Geodesi, Stommätning](#) (HMK-Ge:S), [HMK-Geodesi, Detaljmätning](#) (HMK-Ge:D), [HMK-Databaser](#) (HMK-Da).

Annan inhemsk influens utgörs av till exempel

- Lantmäteriet, LU, KTH och HiG, 2013: [Geodetisk och fotogrammetrisk mätning- och beräkningsteknik](#). (Kompendium).

- SIS, Swedish Standards Institute, 2004. [Samverkande GIS med ISO 19100 – en handbok om tekniskt ramverk för geografisk information.](#)
- SIS-TR 40:2012: [Geografisk information – Tekniskt ramverk – handbok för dataproduktspecifikation](#)

Dessutom är boken

- Lars Harrie, redaktör (2013): [Geografisk informationsbehandling – teori, metoder och tillämpningar](#), upplaga 6. Lund, Studentlitteratur.

en viktig källa för den som vill sätta sig in i detta med geodata i allmänhet och geodatakvalitet i synnerhet. Som svenskspråkig lärobok blir den ett naturligt komplement till HMK, och den harmonierar mycket väl med ovan nämnda dokument och de internationella standarderna.

1.4 Internationella standarder och bidrag

Internationellt finns influenser från framför allt följande dokument:

- 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)
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 19100-serien om geografisk information återfinns på SIS-Stanli:s websida:

<http://www.sis.se/tema/Geodata/geodata-verksamhet/>

Även följande referens är viktig:

- EuroGeographics, *Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies.*
http://www.eurogeographics.org/sites/default/files/Guidelines_ISO_19100_Quality.pdf

1.5 Det norska standardiseringsarbetet

Norge har länge satsat på geodatasamordning/-standardisering.

- **SOSI** (SOSI-standard): *Samordnet Opplegg for Stedfestet Informasjon*, är ett system för standardiserad beskrivning av digitala geodata.
- **Geovekst** är ett geodatasamarbete som syftar till att geografisk information samlas in en gång och ajourhålls av en organisation. Ett samarbetsavtal ingicks 1992 mellan Statens vegvesen, Energiforsyningens Fellesorganisasjon, Kommunenes Sentralforbund, Statens kartverk, Telenor samt Landbruks- og matdepartementet.
- **Norge Digitalt** är ett brett samarbete mellan verksamheter som har ansvar för att ta fram lägesbunden information och/eller som är stora användare av sådan information. Norge digitalt har sin förankring i Stortingsmelding nr. 30 (2002-2003): [Norge digitalt – et felles fundament for verdiskaping](#).
- **FKB** (*Felles KartdataBase*) består av strukturerade vektordata. Data är specificerade i FKB-standarderna FKB-A, FKB-B, FKB-C och FKB-D, som ska täcka behovet av gemensamma geodata.

Relaterade norska standarder är i första hand:

- Statens kartverk, 2001: [Kvalitetssikring av oppmåling, kartlegging og geodata \(Geodatastandard\)](#).
- Statens kartverk, 2007. [Kontroll av geodata](#).
- Statens kartverk, 2009. [Kart og geodata](#).

Dessa är för närvarande föremål för en översyn (remiss under hösten 2014), vilket bland annat innebär att *Kontroll av geodata* och *Geodatastandard* slås ihop till den nya standarden *Geodatakvalitet*.

Den norska standarden "Kontroll av geodata"

"Kontroll av geodata" analyserades i detalj i den första rapporten i HMK-skriftserien *Tekniska rapporter* på: www.lantmateriet.se/hmk under [Aktuella HMK-dokument](#).

Det ter sig rimligt att anpassa kontrollen av datakvalitet i HMK till motsvarande norska dokument – åtminstone som metodbeskrivningar. Det vill säga att kvalitetskontroll av geodata bör kunna göras på samma sätt som i de norska standarderna, utan att det ställs krav på att **alltid** genomföra **samtliga** kontroller.

Det finns flera faktorer som talar för att anamma det norska konceptet:

- Det är beprövat och bygger på internationella ISO-standarder, samt på vedertagna statistiska metoder och delvis på tankegångarna inom GUM (se avsnitt 2.8).
- Det harmonierar med HMK-konceptet, men innehåller användbara utvidgningar.
- Det finns fördelar med att ha en gemensam syn på geodatakvalitet inom Norden, som är en och samma marknad för producenterna; det blir enklare och därmed billigare för alla.

Svensk/norskt samarbete

Sammantaget

- vill vi gärna, genom samarbete, åstadkomma ett gemensamt koncept för datakvalitet och kvalitetskontroll i Norden/Skandinavien
- vill vi försöka sprida detta koncept globalt genom vår medverkan i det internationella standardiseringsarbetet, till exempel inom ISO
- bedömer vi att det som i första hand krävs är språkliga/terminologiska anpassningar.

Eftersom en översyn av de norska standarderna pågår parallellt med HMK-arbetet bedrivs remissarbetet "bilateralt", det vill säga att det ena landet utgör remissinstans för det andra landets dokument. Redan i denna version av vårt dokument finns många influenser från de norska motsvarigheterna, och vi har från svensk sida gjort motsvarande intryck beträffande de norska standarderna.

Några motsvarande kontakter med övriga nordiska länder finns inte för tillfället.

2 Statistik, introduktion

2.1 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 geodatabranschen är även storheten

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

vanligt förekommande. Sambanden mellan dessa statistiska storheter beskrivs i avsnitt 5.3.

2.2 Terminologi

Den terminologiska grunden för kontroll av geodatakvalitet redovisas i Tabell 2.1.

Tabell 2.1. Grundläggande terminologi för kontroll av datakvalitet.

Krav	<i>Specificerat</i> 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.

2.3 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 kvalitetsmått 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.

Kvalitetsmättet måste nämligen vara *signifikant* sämre än kravet för att datamängden ska underkännas. Denna signifikans mäts via toleransen. (Kvalitetskontroll hanteras vidare i kapitel 5.)

2.4 Kravet på signifikans

Signifikans är alltså 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 både från producentens och 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.

2.5 Konfidensintervall

Kvalitetsmättet 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 %, som är det vanligaste värdet och som normalt tillämpas i HMK.

2.6 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.

2.7 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 2.7.a-b vad gäller relationen mellan konfidensintervall och hypotesprövning - eller mellan redovisning och kontroll.

Exempel 2.7.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 2.7.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.

2.8 Mätosäkerhet – GUM

Uppgift om *mätosäkerheten* utgör en integrerad del av lägesbestämningar och andra mätningar. För att skapa en gemensam och bred förståelse bör internationella standarder följas.

I omarbetningen av HMK kommer alla skrivningar som rör noggrannhet och datakvalitet att terminologiskt anpassas till *Guide to the Expression of Uncertainty in Measurement* (GUM). Det centrala begreppet är *mätosäkerhet*.

Arbetet initierades av *Internationella byrån för mått och vikt* (BIPM) och resulterade i ett förstadium till standard år 1980. Översyn av guiden har skett i flera omgångar och under flera huvudmän.

Nuvarande version av dokumentet ([JCGM 100:2008](#)) förvaltas av konsortiet *Joint Committee for Guides in Metrology* (JCGM), där bland annat det internationella standardiseringsorganet ISO ingår.

Geodesin, fotogrammetrin och mätningstekniken har hittills stått utanför detta standardiseringsarbete. I motsats till andra branscher hade geodesin en fungerande hantering av dessa frågor. Till stor del berodde det på att Gauss, som var geodet, hade lagt grunden till analys av mätdata genom innovationerna *normalfördelningen* och *minsta-kvadratmetoden* redan runt sekelskiftet 1700-1800.

I dag finns all anledning att tillämpa standarden även inom svensk mätningsteknik, eftersom enhetlighet och standardisering blir viktigare när gränserna mellan olika branscher och discipliner suddas ut. Ett exempel är GPS, som började som en teknik inom navigation och geodesi men som nu är vardagsteknik.

Dessutom, i vår strävan om "ökad och breddad användning av geodata", kommer med stor sannolikhet nya användare att ha med sig GUM som en naturlig del i bagaget – och då måste vi åtminstone förstå varandra. Därför har vi bestämt att "HMK pratar GUM" vad gäller **kvantitativ** datakvalitet.

Skillnaden gentemot det traditionella synsättet är delvis terminologisk, men det finns även en olikhet i grundfilosofin. Tidigare rörde diskussionen *mätfel* och *felanalys* i stället för *osäkerhet* och *osäkerhetsanalys*. Fel innebär att de gjorda mätningarna relateras till motsvarande **sanna** värden, som i princip aldrig går att hitta.

Osäkerhetsbegreppet utgår endast från **observerbara** data (engelskans *measurands*). För att citera GUM: "Mätosäkerheten är en parameter som är förbunden med mätresultatet och som kännetecknar spridningen av värden som rimligen kan tillskrivas mätstorheten".

Traditionellt har indelningen varit: *slumpmässiga*, *systematiska* och *grova fel*. Men till exempel slumpmässiga fel är ju egentligen inte fel utan naturliga variationer i mätningarna.

Därför tillämpar GUM terminologin

- slumpmässiga avvikelser
- systematiska avvikelser
- grova fel, det vill säga direkta felaktigheter eller "tabbar".

GUM:s grundfilosofi baseras alltså på att slutsatser får dras ur egna mätningar – alternativt med hjälp av osäkerhetsmått som framtagits tidigare (osäkerheten i en viss metod, ett visst instrument etcetera). GUM skiljer därför på bestämning av mätosäkerheten enligt *Typ A* eller *Typ B*:

- **Typ A:** Mätosäkerheten bestäms utifrån det egna mätresultatets variation.
- **Typ B:** Övriga sätt att bestämma mätosäkerheten, till exempel resultat från andra mätningar eller värden tagna från handböcker, kalibreringsbevis etcetera.

Observera att klassificeringen avser sättet att bestämma mätosäkerheten. Osäkerheterna som sådana har inte olika karaktär, och ingen av de två typerna har någon fördel gentemot den andra. Det förekommer även blandningar av de två typerna.

Om mätosäkerheten rör en position i ett referenssystem använder vi i HMK ofta benämningen *lägesosäkerhet* för att det ska bli tydligare. Det beror på att det geografiska läget är så centralt i samband med geodata – i övrigt är hanteringen av de två begreppen lika.

I [HMK - Ordlista och förkortningar](#), kapitel 1, redovisas en förteckning över de vanligaste GUM-termerna och deras motsvarighet i den tidigare – nationella och internationella – terminologin. Mer underlagsmaterial om mätosäkerhet och GUM-standarderna finns på HMK:s hemsida:

www.lantmateriet.se/HMK under *HMK-Referensbibliotek/GUM*.

Originaldokumentet är:

ISO/IEC Guide 98-3:2008. *Uncertainty of Measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM: 1995)*.
www.iso.org/sites/JCGM/GUM-introduction.htm

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.

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

Det finns egentligen två lägen:

- *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 3.

Olika huvudmän befinner sig ofta i olika delar av tabellen. 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 och i detta dokument – ligger på förstagångsinsamlingen. Över tid måste dock ajourhållningen hanteras utförligare – hur den bör utföras och redovisas.

I ajourhållningsskedet kommer liknande 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.

Tabell 3. Typiska skillnader mellan förstagångsuppbyggnaden av en datamängd och dess kontinuerliga ajourhållning därefter – sett ur ett datakvalitetsperspektiv. (DPS = Dataproduktspecifikation, se avsnitt 4.4.)

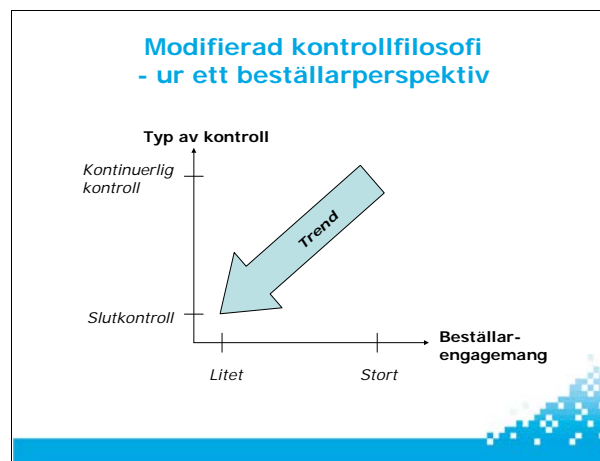
	Förstagångsuppbyggnad	Ajourhållning, kontinuerlig eller periodisk
Aktualitet	Lika för hela datamängden, men beror på underlaget	Beror på angivna ajourhållningsregler enligt DPS
Aktörer	Få aktörer, ofta ett upphandlat uppdrag	Flera aktörer inblandade
Insamlingsmetod	Enhetlig	Varierar med teknikutvecklingen
Krav	Enhetliga och tydliggjorda i DPS	Kan variera över tid; till exempel kan nya användningsområden tillkomma som ställer andra krav
Omfattning	En omfattning 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 kan göras på övergripande nivå	Variationerna över tid innebär att redovisningen måste göras på objektnivå
Typ av kontroll	Kontroll av hela datainsamlingsprocessen	Kontroll av slutresultatet
Underlag	Enhetligt	Varierar över tid; består delvis av objekt som redan finns i datamängden

Om datamängdens användbarhet i befintligt skick inte räcker kan kvalitetsinformationen användas för att bedöma var eventuella kompletteringar behöver göras – alternativt om nykartering krävs.

3.1 Förändrad syn på kvalitetskontroll

Synen på kontroll av datakvalitet har delvis förändrats. 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 större eget ansvar för kontroller under vägen (Figur 3.1). Det förekommer att beställarens slutkontroll upphandlas.



Figur 3.1. Utvecklingen vad gäller beställarkontroll av geodata går 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.

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](#), avsnitt 2.2.

Följande jämförelse är hämtad från laserskanning; ibland bör det nya synsättet väljas, i andra fall är det fortfarande befogat att tillämpa synsättet i de äldre HMK-skrifterna:

- 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 laser-datahanteringen.
- 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.

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.

3.2 HMK-standardnivåer

Standardnivå i HMK definieras som ”rekommendationer för beställarens val av metod/parametrar vid geodatainsamling för visst användningsområde”. HMK-standardnivåerna baseras dels på kraven från beställare, användare, tillämpningar och produkter – dels på de tekniska produktionsmöjligheter som finns i dag.

Standardnivåerna utgör därigenom bryggan mellan beställarens krav och utförarens teknikval. De ger också ett begränsat antal varianter, vilket effektiviserar och standardiserar produktionen och gör datakvaliteten mer homogen. Konceptet fungerar naturligtvis även för egen regi-verksamhet, men också då blir det ofta tydligare om man ser det hela ur ett ”beställarperspektiv”. Kvalitetskontrollen förändras inte nämnvärt om det är det ena eller det andra.

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 för 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 vid geodetisk detaljmätning 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 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 vid geodetisk detaljmätning 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 vid geodetisk detaljmätning och med bilddata från flygplan med en upplösning på decimeternivå eller bättre. Även geodetisk detaljmätning med totalstation och gemensam insamling av bild- och laserdata från helikopter och markfordon förekommer.

HMK-standardnivå 3

HMK-standardnivå 3 används för projektinriktad mätning och kartläggning för projektering, byggande och förvaltning av bebyggelse, vägar och övrig infrastruktur samt för 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) vid geodetisk detaljmätning 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 och markfordon. Även insamling av bilddata med UAV² (*unmanned airborne vehicles*) börjar nu tillämpas.

Sammanfattning

En sammanställning per HMK-standardnivå redovisas i Tabell 3.2.

Tabell 3.2. 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	RTK eller Nätverks-RTK	Totalstation, projektanpassad Nätverks-RTK eller 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

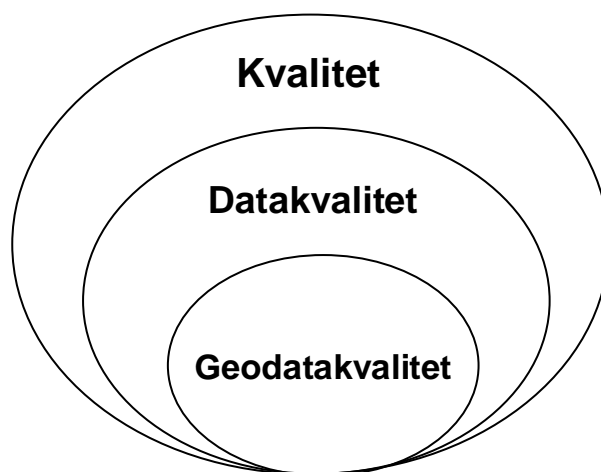
² Benämns även RPA (*Remotely Piloted Aircraft*).

4 Datakvalitet

4.1 Definition av geodatakvalitet

Rent språkligt så försöker vi att i texten precisera kvalitetsterminologin. Det gör vi genom att ersätta den allmänna termen *kvalitet* med *datakvalitet*, eller *geodatakvalitet*, när det är det som avses. För att inte språket ska bli alltför tungt får det dock ibland vara underförstått, och framgå av sammanhanget, vilken typ av kvalitet det är fråga om – i enlighet med hierarkin i Figur 4.1.

Avsteg från preciseringen kan också behöva göras för att terminologin ska överensstämja med, till exempel, officiella standarder.



Figur 4.1. Hierarkin för kvalitetstermer – från det allmänna till det precisa

Geodatakvaliteten redovisas med hjälp av ett antal *kvalitetsparametrar* för en datamängd, se avsnitt 4.5. Kvalitetsredovisningen kan göras på *objektsnivå* eller på en *aggregerad nivå*, det vill säga på detaljnivå eller för hela eller stora delar av datamängden.

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 geodainsamling. I ISO 19158 definieras, förutom datakvalitet enligt ISO 19157, även omfattning, tidplan och kostnad som kvalitetsparametrar vid geodainsamling, se även [HMK-Introduktion](#), avsnitt 2.2.

4.2 Datakvalitetsmodell och informationsmodellering

Användarna av data har många och ibland skilda behov. Datakvalitetsdokumentationen används för att bedöma om en viss datamängd uppfyller en viss användargrups behov.

Några exempel på denna typ av dokumentation beträffande en *geodatamängd* 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 (se nästa avsnitt).

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

4.3 Metadata

Det finns ett stort behov av översiktlig information om databasers innehåll och datakvalitet. För sådana uppgifter, ofta lagrade i särskilda databaser, används termen *metadata*.

Metadata betyder rent språkligt "data om data". De beskriver datamängdens innehåll och struktur - vanligen ur ett speciellt perspektiv. Syftet är att kunna söka, hitta och utvärdera data och tjänster (namn på produkt, vem som levererar, restriktioner etcetera).

De ska bland annat visa graden av överensstämmelse mellan dataproduktens totala, faktiska beskaffenhet och de intentioner som anges i *dataproduktspecifikationen*. I metadata redovisas därför information om eventuella kvalitetsutvärderingar - men även om sådant som tillgänglighet, rättigheter, ajourhållningsregler, pris etcetera.

En ny internationell metadatastandard är nyligen fastställd (SS-EN ISO/19115-1:2014). På www.geodata.se finns även en [nationell metadataprofil för geodata](#), som utgör specifikation och vägledning för att ta fram enhetliga metadata. Den kommer att ses över för att passa till nyttgåvan av standarden.

Metadataprofilen tillgodoser kraven från Inspire och är anpassad för bland annat dataproducenter, användare och systemleverantörer som vill publicera metadata i Geodataportalen.

I förhållande till i dag finns ett generellt behov av metadata på objektnivå.

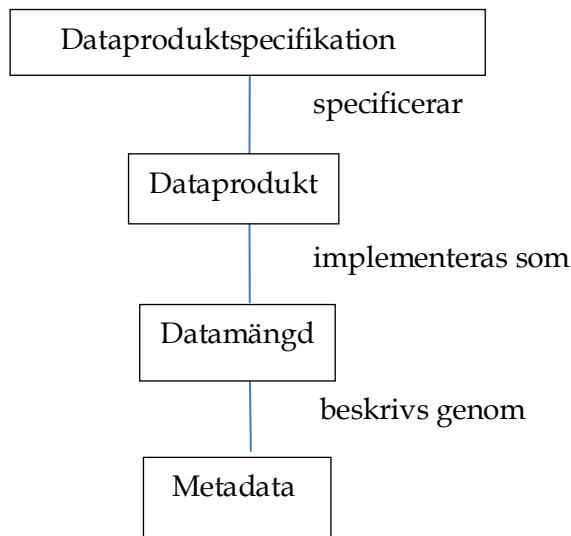
4.4 Dataproduktspecifikationen

Grunden för redovisningen av datakvalitet är *dataproduktspecifikationen*. En dataproduktspecifikation (DPS) upprättas för varje dataprodukt (ett geografiskt tema eller en karta) och utgör en beskrivning av till exempel innehåll, datakvalitet och metadata.

Med begreppet *dataprodukt* avses såväl leveransen från en dataproducent till en dataförvaltare som innehållet i en dataförvaltares databas eller en leverans från dataförvaltare till en användare (se Figur 5.1). Dataproduktspecifikationerna följer ISO 19131.

De kvalitetsuppgifter som finns i dataproduktspecifikationerna anger de krav som är ställda på dataprodukterna. Utifrån dessa krav kan man ställa vissa förväntningar på produkternas användbarhet. Datamängden ska uppfylla de krav specifikationen anger. Mottagaren kan vid leveransen kontrollera om kraven är uppfyllda.

Figur 4.4 visar sambanden mellan dataproduktspecifikation, dataprodukt, datamängd och metadata.



Figur 4.4: Relationer mellan dataproduktspecifikation och metadata (ISO 19131).

Se vidare i SIS-TR 40:2012 *Handbok för dataproduktspecifikation*. En checklista för innehåll i och upprättande av sådana specifikationer återfinns i Bilaga B.1.

[Svensk geoprocess](#) jobbar med [enhetliga dataproduktspecifikationer](#) för nio grundläggande datateman: *stompunkter, flygbilder/ortofoton, laserdata/höjdmmodell, marktäcke/markanvändning, hydrografi, kommunikation, markdetaljer/tekniska anläggningar, byggnader samt adresser*.

4.5 Kvalitetsteman och kvalitetsparametrar

De kvalitetsteman för datakvalitet som beskrivs i ISO 19157 är *fullständighet, logisk konsistens, lägesnoggrannhet, tematisk noggrannhet, temporal noggrannhet* samt *användbarhet*. I HMK används termerna *mät- eller lägesosäkerhet* i stället för *noggrannhet*, enligt GUM, om det inte handlar om direkta citat från standarder.

Primära parametrar

Datakvalitet kan beskrivas med hjälp av de kvalitetsteman och kvalitetsparametrar som redovisas i Tabell 4.5.

Därutöver brukar man ibland redovisa *syfte, spårbarhet* samt *aktualitet*. Dessa kvalitetsteman finns omnämnda i standardiseringsarbetet men har inte formellt inkluderats i standarderna. Antingen inkluderas redovisningen under "Användbarhet", eller också krävs en utvidgning i förhållande till ISO-standarderna. Formen blir av typen fri-text, så länge inte kvalitetsparametrar/kvalitetsmått har tagits fram.

Tabell 4.5. Primära kvalitetsteman och kvalitetsparametrar för datakvalitet.

Fullständighet	Datamängdens innehållsmässiga överensstämmelse med dataproductspecifikationen; 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 eller relationer (till exempel sammanhängande nätverk och slutna ytor).
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>konceptuell konsistens</i> (giltiga kombinationer av värden, som stämmer med informationsmodellen) - <i>domänkonsistens</i> (giltig värdemängd, dvs. tillåtna värden) - <i>formatkonsistens</i> (korrekt format) - <i>topologisk konsistens</i> (följer angivna topologiska regler).
Lägesnoggrannhet	Noggrannhet/mätosäkerhet i position.
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>absolut lägesnoggrannhet</i> (i förhållande till referenssystemet) - <i>relativ lägesnoggrannhet</i> (i förhållande till närliggande objekt) - <i>lägesnoggrannhet hos rasterdata</i>.
Tematisk noggrannhet	Noggrannhet hos kvantitativa attribut och huruvida kvalitativa attribut och klassificeringar är korrekta.
	Kvalitetsparametrar: <ul style="list-style-type: none"> - <i>klassificeringsnoggrannhet</i> (korrekthet beträffande objekttyp) - <i>tematisk noggrannhet, kvalitativa attribut</i> (icke mätbara) - <i>tematisk noggrannhet, kvantitativa attribut</i> (mätbara).
Temporal noggrannhet	Noggrannhet/osäkerhet för temporala (tidsmässiga) attribut och temporala förhållanden mellan objekt.

	<p>Kvalitetsparametrar:</p> <ul style="list-style-type: none"> - <i>tidsnoggrannhet</i> (noggrannhet 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.
	<p>Kvalitetsparametrar:</p> <p>Om övriga kvalitetsparametrar inte uttrycker en speciell datakvalitetsegenskap tillräckligt bra, kan användbarhet tillämpas. Den är fri till formatet.</p>

Syfte

Att formulera ett tydligt *syfte* med en dataproduct har många fördelar – såväl för den som ansvarar för geodata och den som ska producera eller förädla data som 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 dataproductspecifikationen. Genom att definiera syftet kan man också revidera detta om det efter hand uppstår en efterfrågan av en från början oförutsedd användning.

Spårbarhet

I en kvalitetsmedveten produktion av geodata är det viktigt att även redovisa datas *spårbarhet* bakåt i produktionskedjan; synonymmer 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 inte enbart till att man korrigerar avvikande data utan även till att man kan spåra orsaken och åtgärda den.

- Med kunskap om tillkomsthistoria och tidigare tillämpning kan en användare även bedöma datamängdens användbarhet för ett visst syfte.

Data och information om spårbarheten i sig bör lagras i databas-tillämpningens eller informationssystemets metadata.

Aktualitet

Aktualitet finns inte med i datakvalitetsstandarden ISO 19157 som en egen kvalitetsparameter, även om det är en datakvalitetssegenskap som kunder ofta efterfrågar.

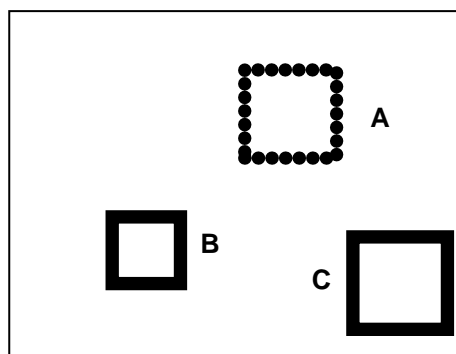
Aktualitet är ett sätt att i en producent-/leverantörsroll mäta och löpande redovisa hur data av olika typer åldras och förlorar i informationsvärde för avsedd tillämpning. Aktualiteten skattas lämpligen med hjälp av den tidpunkt då objektet senast, genom kontroll, konstaterades vara korrekt redovisat.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

Aktualitet kan också relateras till *ledtider för ajourhållning*. Vid till exempel kontinuerlig ajourhållning, om det uppstår långa ledtider från det att en förändring har skett tills det att data är uppdaterade i databasen, blir konsekvensen att datamängden inte är helt aktuell.

4.6 Relationer mellan olika kvalitetsparametrar

Om man studerar varje kvalitetstema eller kvalitetsparameter för sig så ser modellen enligt Tabell 4.5 ganska tydlig ut, men om man inbördes börjar relatera parametrarna till varandra så ter sig inte allt lika självklart. Låt oss titta på två enkla men realistiska exempel.

Exempel 4.6.a: I dataproduktspecifikationen anges att byggnader > 20 kvadratmeter ska ingå. Ajourhållningscykeln för byggnader är 1 år. I Figur 4.6.a redovisas ett utsnitt.

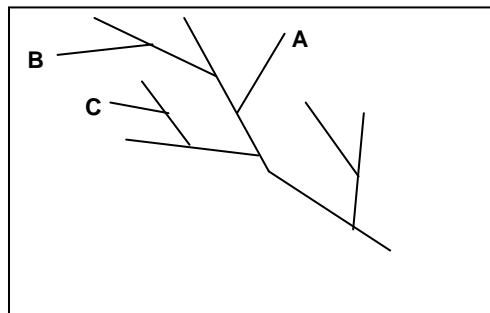


Figur 4.6.a. Kontroll av objekttypen "Byggnad".

Några reflektioner:

- A finns inte i databasen trots att byggnaden uppfyller specifikationens krav. Den uppfördes för cirka ett halvår sedan. Brist i fullständigheten eller i aktualiteten - eller helt korrekt eftersom endast årlig ajourhållning garanteras?
- B är uppmätt till 15 kvadratmeter. Den är med men borde inte vara det. Övertalighet inom objekttypen "byggnad" (korrekt area) eller brister i noggrannheten för kvantitativa attribut (fel area)?
- C är en stor vedtrave med plåttak. Den är tillräckligt stor men är inte en byggnad. Övertalighet inom objekttypen "byggnad" eller brister i klassificeringsnoggrannheten (fel objekttyp)?

Exempel 4.6.b: Ett avrinningsområde ska översiktligt representeras av ett topologiskt nätverk bestående av vattendrag bredare än 2 meter, se Figur 4.6.b.



Figur 4.6.b. Kontroll av ett avrinningsområde i form av ett topologiskt nätverk

Reflektioner:

- A finns i databasen trots att det vattendraget är totalt uttorakat. Är det brister i klassificeringsnoggrannheten eller brister i logisk konsistens att A finns med? Objektet får felaktigt en topologisk relation till övriga vattendrag. Det innebär även en övertalighet vad gäller fullständigheten.
- Samma sak beträffande B, fast tvärtom. B uppfyller kraven enligt specifikationen men saknas i basen. Objektet borde finnas med och ha en topologisk relation - men har det inte. Det blir därför även en brist vad gäller fullständigheten.
- C saknas också, på grund av att dess bredd felaktigt har redovisats som < 2 meter medan den i verkligheten är > 2 meter. Där är orsaken primärt brister i klassningsnoggrannheten.

Dessa frågeställningar tas även upp i SIS-TR40:2012, bilaga B, samt i boken "Geografisk informationsbehandling" (Harrie et.al., 2013).

4.7 Kvalitetsparametrar vs. kvalitetsmått

Kvalitetsmått är alltså begreppet och verktyget för att specificera hur datakvalitet ska mätas och dokumenteras. Det är också det centrala begreppet för att ställa krav på datakvaliteten. Avsikten med ett kvalitetsmått är att specificera hur man genom att använda en speciell mät- och/eller beräkningsmetod kommer fram till ett tal som beskriver hur bra en del av en datamängd är i ett visst avseende.

Varje kvalitetsmått "mäter" en viss bestämd kvalitetsegenskap i datamängden, och genom att jämföra mätvärdet med kraven i specifikationen kan man avgöra om datamängden uppfyller kraven eller ej.

Genom att studera utfallet för samtliga kvalitetsmått kan man få en samlad bild av geodatamängdens datakvalitet.

Intuitiv tolkning av "noggrannhet"

Vad betyder utsagan: *Noggrannheten i en orienteringskarta är 7 meter?*

Det är ganska oklart, och beroende på vem som gör uttalandet så varierar sannolikt innebörden. Vi måste bli tydligare när vi uttalar oss om noggrannhet – egentligen *lägesosäkerhet* – och datakvalitet överhuvudtaget!

Det räcker inte med ett kvalitetstema eller en kvalitetsparameter utan vi måste framför allt tala om vilket kvalitetsmått vi har använt och hur vi har bestämt kvalitetsmåttets värde.

Det vill säga, de frågor man alltid bör ställa sig när man får kvalitetsuppgifter är:

- Vilket kvalitetsmått har Du använt?
- Hur har Du bestämt dess värde?

Allting utgår från kraven i dataproduktspecifikationen.

Exempel 4.7.a: Lägesosäkerheten i en orienteringskarta är 7 meter.

Vilket kvalitetsmått har Du använt? Standardosäkerheten i plan (punktmedelfel) i förhållande till referenssystemet, det vill säga absolut lägesosäkerhet.

Hur har Du bestämt dess värde? Genom kontrollmätning av tydliga detaljer med Nätverks-RTK.

Kvalitetsmått för andra kvalitetsparametrar

Hur mäter man då på motsvarande sätt andra aspekter på datakvalitet? Det vill säga, vilka kvalitetsmått kan kopplas till övriga kvalitetsparametrar?

Boken "Geografisk informationsbehandling" (Harrie et.al., 2013) ger en bra grund till sådana kvalitetskontroller i kapitel 10. Där definieras bland annat olika mått inom andra kvalitetsteman än lägesnoggrannhet/mätosäkerhet.

En av de grundläggande termerna är *förväxlingsmatriser*, ur vilka diverse kvalitetsmått kan härledas och beräknas, framför allt vad gäller klassningsnoggrannhet/-osäkerhet. Vi går inte här in i denna fråga på djupet utan nöjer oss med några enkla exempel, 4.7.b och 4.7.c.

Exempel 4.7.b: Fullständigheten i en digital primärkarta är 95 %.

Vilket kvalitetsmått har Du använt? *Andelen objekt som finns med och som uppfyller dataproduktspecifikationens definition.*

Hur har Du bestämt dess värde? *Genom fältkontroll av ett stickprov.*

Exempel 4.6.c: Klassificeringsnoggrannheten i databasen är 97 %.

Vilket kvalitetsmått har Du använt? *Andelen objekt som är korrekt kodade enligt dataproduktspecifikationens definition.*

Hur har Du bestämt dess värde? *Genom kontroll i nyligen tagna, storskaliga flygbilder.*

Beskrivning enligt ISO 19157

Enligt ISO 19157 beskrivs kvalitetsmått så som beskrivs i Tabell 4.7.

Tabell 4.7. Beskrivning av kvalitetsmått (utdrag ur ISO 19157).

Egenskap	Beteckning i ISO 19157	Förklaring
namn	name	kvalitetsmåttets benämning
definition	definition	kortfattad förklaring
beskrivning	description example	förklaring av kvalitetsmåttet; exempel på dess användning
källreferens	source reference	referens till en mer utförlig dokumentation

Ett norskt register för kvalitetsmått redovisas i Bilaga D.

4.8 Intervallskala för lägesosäkerhet

Lägesosäkerheten i geometriska data redovisas primärt i form av standardosäkerhet (medelfel). 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 lägesnoggrannhet/mätosäkerhet, avseende standardavvikelse (medelfel), 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 4.8: I Stockholm stad har man definierat följande krav på fältmätningen 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 3.2.

4.9 Begreppet kontrollerbarhet

I de äldre HMK-skrifterna används begreppet *kontrollerbarhet* i stor utsträckning, särskilt i [HMK-Stommätning](#). Kontrollerbarheten anger hur pass svårt/enkelt det är att hitta grova fel i ett mätmaterial. Även om vi försöker undvika begreppet "fel" så förtjänar eliminering av just denna typ av felaktig hantering (till exempel avläsningsfel, skrivfel och andra slarvfel) sin plats i presentationen.

Kontrollerbarheten – vanligen i ett geodetiskt nät – mäts med det så kallade *k-talet*, som definieras som:

$$k = \frac{n-m}{n} = \text{antalet överbestämningar/antalet mätningar}$$

där m = antalet obekanta i den aktuella beräkningen.

k -talet ger en uppfattning av hur stor del av ett grovt fel som är synligt i motsvarande förbättring efter en minsta-kvadratutjämnning. Vid ett k -tal på 0,5 "syns" cirka 50 % av ett eventuellt grovt fel. Ett decimeterfel ger därför upphov till en förbättring på runt 5 centimeter – resten "smetas ut" genom utjämningsberäkningen och ger en större osäkerhet i beräknade koordinater, höjder etcetera.

Exempel 4.9.a: Beträffande k -tal i traditionella geodetiska nät gäller:

- I triangelnät och vid fri station bör $k \geq 0,5$ eftersträvas.
- Typiska värden för höjdnät (avvägningsnät) är $k \approx 0,3$.
- I polygontåg med många punkter är $k \approx 0,1$ vanligt.

Exempel 4.9.b:

- Ett avstånd mäts med den kända standardosäkerheten 50 mm. Samma avstånd mäts, av en annan mätare, med standardosäkerheten 25 mm.
- I det första fallet görs 4 upprepade mätningar, i det andra bara en mätning.
- I det första fallet blir standardosäkerheten för medelvärdet av de 4 mätningarna $= s/\sqrt{4} = 50/2 = 25$ mm. Det är samma värde som den enda mätningen i det andra fallet.
- I det första fallet blir kontrollerbarheten $k = 3/4 = 75\%$ (antalet överbestämningar dividerat med antalet mätningar). I det andra fallet får vi $k = 0/1 = 0\%$.

Utan överbestämningar erhålls ingen kontroll på avvikelser och grova fel!

Tillförlitlighet är en med kontrollerbarhet närbesläktad term, som beskrivs i HMK-Stommätning. Båda begreppen kan ses som en sorts kvalitetsparametrar till *Lägesosäkerhet*.

5 Kontroll av geodata

För att säkerställa viss kvalitet och för att redovisa faktisk kvalitet för en dataprodukt måste data genomgå kvalitetskontroll.

5.1 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 betydligt enklare att utföra – men resultatet kan naturligtvis 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 med hjälp av lämplig programvara – till exempel ytbildning för att testa topologi (logisk konsistens). Full kontroll används också när antalet förekomster av en objekttyp är så litet att stickprovskontroll inte är tillämpbar.

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

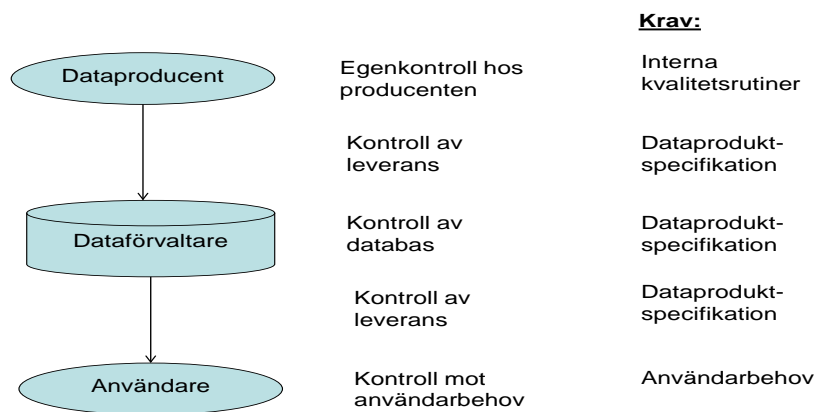
Egenkontroll avser den kontroll som utföraren av arbetet gör. Resultatet redovisas till beställaren, men denne kan även göra egna kontroller, *beställarens kontroll*. Ofta upprättas särskilda *kontrollprogram* som en del av ett mätuppdrag. Detta gäller även egen regi-verksamhet, och i båda fallen ligger dataproduktspecifikationen till grund.

Mycket av kvalitetskontrollen – av kvantitativa kvalitetsparametrar – bygger på *överbestämningar*, det vill säga fler mätningar än vad som krävs för att få ett resultat – till exempel dubbelmätning av längder, återbesök på samma detaljpunkt, inmätning från två höjdfixar samt inmätning av redan kända punkter.

Vid detaljmätning är vanligen den lokala överensstämelsen viktigast – till exempel att närliggande detaljer ligger rätt i förhållande till varandra, att hus har rätt storlek och är rektangulära. Den lokala överensstämelsen kan kontrolleras genom separat längdmätning (avstånd mellan byggelement, fasadmått, jämförelse mellan de båda diagonalerna i en utsatt byggnad etcetera).

I Figur 5.1 sätts olika typer av kontroll in i flödet

Producent → *Förvaltare* → *Användare*.



Figur 5.1. Olika typer av kontroller på vägen från producent till användare.

5.2 HMK:s principer för toleranser (felgränser)

Hela mätosäkerhetskonceptet bygger på redovisning av **faktisk** mätosäkerhet. I många branscher regleras dock mätkvaliteten i form av fastställda *toleranser*, det vill säga gränsvärden för avvikelser, osäkerhetsmått etcetera som inte får överskridas.

HMK innehåller sådana toleranser för olika typer av mätningar. I [HMK-Detaljmätning](#), bilaga F, finns till exempel toleranser för detaljmätning och i [HMK-Stommätning](#), finns toleranser för stommätning – fast där benämns toleranserna *felgränser*, i enlighet med den äldre terminologin.

Nedan följer en moderniserad, beskrivning av dessa principer – i enlighet med terminologin i avsnitt 2.8:

Gränsvärdena för avvikelser mellan upprepade mätningar, förbättringar och beräknade osäkerhetsmått vid utjämning samt kontrollmätningar av slutresultatet benämns toleranser eller felgränser. Den senare har hittills varit den vanligaste termen inom geodesin, men en svängning mot toleranser kan idag skönjas.

Standardosäkerhet, som är det normala osäkerhetsmålet, utnyttjas för att ställa upp toleranser. För avvikelser, till exempel mellan dubbelbestämningar eller mellan ursprunglig mätning och kontrollmätning, tillämpas här genomgående toleransen "två gånger standardosäkerheten" (*utvidgad mätosäkerhet*, 2σ).

Denna princip bygger på toleranser konstruerade så att en standardosäkerhet, som tar hänsyn till osäkerheten i såväl ursprungs- som kontrollmätningen, beräknas och multipliceras med täckningsfaktorn två.

Detta motsvarar 95 % täckningsgrad, och innebär att i normalfallet bör 95 % av alla mätningar eller punktbestämningar klara kravet. Större avvikelser är alltså så osannolika att de bör betraktas som en indikation på att kraven på mätosäkerhet inte uppfylls.

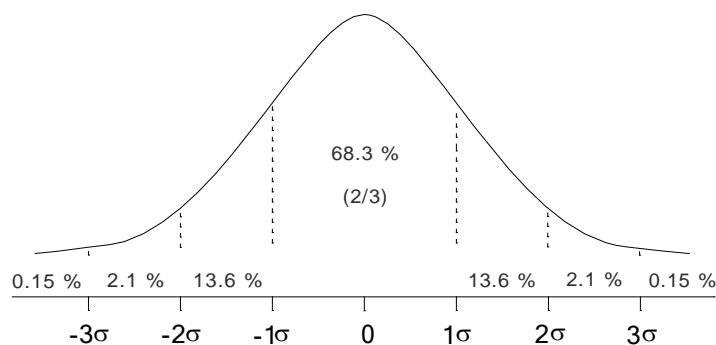
Även för standardosäkerheter beräknade ur mätmaterialet kan toleranser ställas upp. Dessa konstrueras på ett likartat sätt, för att åstadkomma en täckningsgrad på 95 %.

[HMK-Stommätning](#) laborerar med tre nivåer när det gäller osäkerhet. Det bygger på antaganden om standardosäkerheten (σ) och den teoretiska normalfördelningen (Figur 5.2), som säger att:

$\pm 1\sigma$ har täckningsgraden 68,3 %

$\pm 2\sigma$ har täckningsgraden 95,5 %

$\pm 3\sigma$ har täckningsgraden 99,7 %



Figur 5.2. Normalfördelningens frekvensfunktion indelad i 1, 2 respektive 3σ .

Nivåerna tillämpas på följande sätt:

- I. 1σ -gränser används för ett test av att antagandet om normalfördelning är korrekt – ett **fördelningstest**. 2/3 av mätmaterialet bör ha avvikelser som är mindre än detta värde.
- II. 2σ används som **varningsgräns**. Om avvikelserna överskrider denna gräns bör den bakomliggande orsaken analyseras.
- III. Avvikelse större än 3σ betraktas som "grova fel". Därför är 3σ -gränsen att betrakta som en ren **kassationsgräns** och om-mätning krävs.

Exempel på analys av kontrollmätning enligt dessa principer finns i bilagorna A.2-A.4. Tabell 5.2 är hämtad från [HMK-Stommätning](#) – mer för att åskådliggöra strukturen än att lägga fast toleranserna.

Tabell 5.2. Toleranser (felgränser) för slutningsfel i anslutningsnät i höjd; tåglängden anges i kilometer. Från HMK-Stommätning, bilaga A, tabell A.13.

Typ av tåg	Maximalt slutningsfel (mm)		
	I	II	III
Enkeltåg mellan kända punkter	-	$4\sqrt{L}$	$6\sqrt{L}$
Tåg ingående i höjdnät	$1\sqrt{L}$	$2\sqrt{L}$	$3\sqrt{L}$

Toleranser i stället för felgränser

I HMK används alltså numera begreppet toleranser i stället för det "ålderdomliga" felgränser.

Den överordnade termen är:

- Tolerans, som antingen kan vara ett krav på den maximalt tillåtna avvikelsen 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 .

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.

Bakomliggande filosofi

Principen i HMK har genomgående varit att varje delmoment kontrolleras, inte bara slutprodukten. Det finns flera anledningar till det – även om synsättet har luckrats upp (se avsnitt 3.1):

- Det blir enklare att lokalisera orsaken om kontrollen utförs tidigt, och i de flesta fall är det effektivare och billigare att omedelbart göra en nödvändig korrigering.
- En sådan kontroll kan ge underlag för att förbättra mätprocessen i den del som ger ett negativt utslag, för att framtida problem ska undvikas.

Vidare ska toleranserna kopplas till den mätmetod som väljs (mät-toleranser), inte till de krav som ställs på slutprodukten (produkt-toleranser). 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.

Om mätningarna kontrolleras mot toleranser som är generösa i förhållande till mätmetodens osäkerhet, kan besynnerliga effekter uppstå om grova fel förekommer. Grova fel och systematik vill vi helst slippa. Det är en effektiv hantering av slumpmässiga avvikelser som till exempel minsta-kvadratutjämning syftar till.

Det medför att det går 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.

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.

5.3 Systematisk avvikelse, standardavvikelse, RMS och grova fel

Den del där geodatastandarderna bedöms vara förhållandevis outvecklade är de delar som berör noggrannhet/mätosäkerhet. De verkar inte heller följa beprövad praxis inom statistiken. Detta gäller främst lägesosäkerhet (lägesnoggrannhet) men även noggrannhet/mätosäkerhet i de attributdata som mäts.

Bristerna består i huvudsak av följande:

- Kontrollerna är fokuserade på medelfel (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 norska standarden "Kontroll av geodata" på ett mer korrekt sätt. Flera kompletteringar har därför gjorts i Bilaga A, för att åtgärda denna brist. Detta gäller såväl lägesosäkerheten som noggrannheten/mätosäkerheten i (kvantitativa) attributdata.

Systematisk avvikelse, standardavvikelse och RMS (se avsnitt 2.1) beräknas ofta i en och samma operation för att beskriva lägesosäkerhet. I en dimension - till exempel beräkning av höjder - råder följande formelsamband mellan medelvärde (systematisk avvikelse), standardavvikelse och RMS:

$$\bar{x}^2 + s^2 \frac{(n-1)}{n} = RMS^2$$

det vill säga, approximativt för någorlunda stora n (antalet analyserade punkter)

$$RMS^2 = s^2 + \bar{x}^2$$

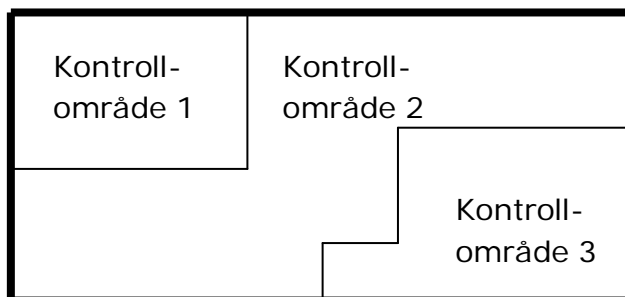
Det är därför naturligt att beskriva samtliga noggrannhetsaspekter i ett sammanhang, eftersom s och \bar{x} var för sig ger en ofullständig information om mätosäkerheten.

I en sådan redovisning bör även ingå lokalisering av ev. *grova fel*. Och man får inte ta bort grova fel ur den fortsatta hanteringen - till exempel vid beräkning av systematik eller standardosäkerhet/RMS - utan att rapportera deras förekomst!

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

- 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.

5.4 Stickprov och kontrollområden



Figur 5.4. En indelning i tre kontrollområden som tillsammans täcker kartläggingsområdet. (Fritt efter nya norska standarden Geodatakvalitet)

Som utgångspunkt för stickprov delas datamängden geografiskt in i *kontrollområden*, på ett sådant sätt att de tillsammans (vanligtvis) täcker hela kartläggingsområdet, se Figur 5.4.

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 datamängden i det kontrollområdet. Hur en sådan brist åtgärdas är något som måste avtalas mellan beställare och utförare. Minst två objekttyper bör kontrolleras.

Tabell 5.4. Stickprovsstorlekar för kvalitativ och kvantitativ kvalitetskontroll. (Identisk med nya norska standarden Geodatakvalitet)

Antal förekomster av objekttypen 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 5.4. Se vidare tillämpningsexemplet i Bilaga C.

6 Läs mer

Utöver de källor och referenser som anges i kapitel 1 bör tilläggas:

Cederholm, T & Persson, C-G, 1989. *Standardiseringsverksamheten i Sverige inom GIS-området – terminologifrågor, informationsstrukturering och kvalitetsmärkning*. Ingår i ULI-rapport 1989:4 "Geografiska informationssystem – föredrag vid ULIs utbildnings- och informationsdagar 1989" och SINUS 1990:2.

Devillers, D & Jeansoulin, R (editors), 2006. *Fundamentals of spatial data quality*. ISTE Ltd, Storbritannien och USA.

Devillers, R, Stein, A, Bédard, Y, Chrisman, N, Fisher, P & Shi, W, 2010. *Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities*. Transactions in GIS, 14(4): 387-400.

Jansson, P & Persson, C-G, 2013. *Analys av den norska standarden "Kontroll av geodata"*. Teknisk rapport 2013:1. Lantmäteriet/HMK.

Persson, C-G, Rost, H, Lithén, T, 2014. *Kontroll av lägesosäkerheten i laserdata*. Teknisk rapport 2014:1. Lantmäteriet/HMK.

Mer information om datakvalitet finns på HMK:s hemsida:

www.lantmateriet.se/HMK under [HMK-Referensbibliotek/Kvalitet](#).

Se särskilt *Övriga hänvisningar*.

A Exempel på kontrollmätningar och toleranser

Här redovisas några exempel på hur kontrollmätningar kan hantearas. 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 5.3). 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) ges en bra redovisning av sådan kvalitetskontroll i dess Kapitel 10.

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** avvikelse från kravet (se avsnitt 2.3). 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 **specifiserad datakvalitet** så finns det en risk att data som uppfyller kraven ändå förkastas.

Exempel A.1: I en studie av NNH (Persson, Rost, Lithén, 2014) visas att standardosäkerheten i höjd är ca. 50 mm. För att lägesosäkerheten i ett skanningområde ska anses **signifikant** avvika från detta värde – med sannolikheten 95 % – så måste avvikelserna (RMS) i kontrolllytorna i genomsnitt överskrida 80 mm, dvs.

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

Att kräva att alla RMS-värden är ≤ 50 mm skulle förkasta cirka hälften av skanningområdena (46 % i nämnda studie), i onödan. Medelvärdet av alla kontrolllytors RMS är just 50 mm. Och att säga att standardosäkerheten i NNH 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 5.2 (n = antalet kontrollmätningar, ε_i är av-

vikelsen 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 $\sigma \cdot (0,96 + n^{-0.4})$, 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 medelavvikelse³ $\bar{\Delta}_R = \sqrt{\bar{\varepsilon}_N^2 + \bar{\varepsilon}_E^2} \leq \frac{2\sigma}{\sqrt{n}} = \frac{2 \cdot 15}{\sqrt{20}} = 7$ 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 täckningsgrad på 95 % ger de täckningsfaktorer som redovisas i Tabell A.2.

³ 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 – ursprungsmätning)⁴ som redovisas i Tabell A.3.a.

⁴ Denna teckenkonvention ger direkt rätt tecken på den korrektion vi räknar fram. Ett fel beräknas tvärtom (ursprungsmätning – kontrollmätning)

Tabell A.3.a. Kontroll av utsättning. Avvikelser 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 av att utsättningen uppfyller kraven på 1, 2 respektive 3σ enligt HMK samt att standardosäkerheten (σ) inte överskrider dataproductspecifikationens angivna krav på 10 millimeter. Kontrollmetoden kan för enkelhets skull betraktas som felfri.

Lösning

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

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

Intervall	Avvikelser i intervallet	Antal
Avvikelser $\leq 1\sigma$	3, 2, -8, 7, 8, -2, 2, 1, 7, -7, 2, -1, -4	13 st.
$1\sigma < \text{Avvikelser} \leq 2\sigma$	14, 16, -12, -11, -19, -18	6 st.
$2\sigma < \text{Avvikelser} \leq 3\sigma$	26	1 st.
Avvikelser $> 3\sigma$	-	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 (%)
Avvikelser $\leq 1\sigma$	67	65
Avvikelser $\leq 2\sigma$	95	95
Avvikelser $\leq 3\sigma$	100	100
Avvikelser $> 3\sigma$	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. Då beräknas 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)⁵. Det innebär att det beräknade RMS klarar toleransen för standardosäkerheten.

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 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. Avvikelser 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.

⁵ Gränsvärdet kan även hämtas från Tabell A.5.b. Den tabellen ger i stället värdet $10 \cdot 1,27 = 12,7$ mm, vilket beror på att formeln ovan bara är en (god) approximation.

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 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 höjdskiftet 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 höjdskift – som ska tillämpas generellt, även på andra mätningar – bör proceduren upprepas eller 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, mätosäkerhet och systematik i 1 D, 2D och 3D

Här förs ett resonemang om 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 konfidenegrad än det förenklade synsättet i Bilaga A.2.

Flera av de angivna värdena har verifierats genom Monte Carlo-simulering (Jansson & Persson, 2013).

Grova fel

3σ - där σ är den **specificerade** standardosäkerheten - används som gräns för grova fel. I 1D motsvarar det konfidensgraden 99,73 %. Denna konfidensgrad i två och tre dimensioner ger $2,43\sigma$ i 2D och $2,17\sigma$ 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 3σ i 1D.

1D	3σ	→	99,73 %
2D	$2,43\sigma$	←	99,73 %
3D	$2,17\sigma$	←	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 % konfidensgrad 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 3σ) upplagd - mest på grund av sin enkelhet, se avsnitt 5.2.

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. Konfidensgrad 95 %. (Identisk med den kommande 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. Konfidensgrad 95 %. (Identisk med den kommande 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 och fullständighet/objektklassificering

Tabell A.6 gäller för kontroll av grova fel, samt för test av fullständighet och objektklassificering, ur stickprov. Kontrollen går ut på att jämföra antalet hittade fel med tabellvärdena för antalet tillåtna fel. 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 grova fel är mindre än 5 %. Det vill säga, vår tolerans 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 grova fel ska vara **mindre** än tabellvärdena blir i realiteten kraven strängare än binomialfördelningens teoretiska värden för angivna konfidensgrader.

(Siffror markerade med **gult** är subjektivt satta.)

Tabell A.6. Kontroll av grova fel och fullständighet/objektklassificering ur stickprov. Antalet grova fel eller fel redovisade objekt ska vara mindre än tabellvärdena. Vid kontroll av fullständighet inkluderas såväl övertalighet (för många objekt) som brist (för få objekt) – allt enligt dataproduktspecifikationen. (Identisk med nya norska standarden Geodatakvalitet)

Antal förekomster av objekttypen i kontrollområdet		Stickprovsstorlek (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 5.4 för kvalitativ kontroll. Ett exempel på deras tillämpning finns i Bilaga C.

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 geodata-aktörer 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*. För fördjupning hänvisas till originaldokumentet. Standarden beskriver delarna som ingående klasser i en databas, 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 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 Omfattning för specifikationen		Anger dataproduktspecifikationens avgränsning.
2 Översikt*		Sammanfattande obligatoriskt avsnitt för dataproduktspecifikationen. Rekommendation är att använda underrubrikerna.
2.1 Information om specifikationen*	Titel Datum Ansvarig part Språk Ämnesområde Syfte	Information om skapandet av dataproduktspecifikationen, viss information kan stå i sidhuvud/sidfot.

2.1.1 Termer och definitioner*		Termer och dess definitioner, kan ersättas med en länk till källan.
2.1.2 Förkortningar*		Förkortningar och dess fulla betydelser.
2.2 Sammanfattning av dataprodukten*	Namn* Förkortning* Beskrivning av innehåll* 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*		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*	Identifiering* Hierarkisk nivå* Namn på hierarkisk nivå* Beskrivning av hierarkisk nivå* Utsträckning* Yttäcken* Ingående omfattningar	Beskrivning av omfattning av hela datamängden. Hierarkisk nivå enligt metadatastandard (ISO 19115), till exempel serie, datamängd och objekttyp. Yttäcken är till exempel raster, grid och TIN.
3.2 Delomfattning	Identifiering* Hierarkisk nivå* Namn på hierarkisk nivå* Förkortning Beskrivning av hierarkisk nivå* Utsträckning* Yttäcken* Ingående omfattningar Överordnad omfattning*	Beskrivning av en del av datamängden. Kan göras i det avsnitt som berörs av omfattningen. Då kan identifiering utelämnas.

4 Identifiering*	Titel* Alternativ titel Versionsnummer Sammanfattning* Syfte Ämnesområde* Metod för rumslig representation Rumslig upplösning Geografisk utsträckning* 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*	Beskrivande text* Applikationsschema Objekttypskatalog* Omfattning*	Utförlig beskrivning av innehållet och strukturen i dataprodukten, där applikationsschema (ISO 19109) och objekttypskatalog (ISO 19110) kan levereras som bilaga eller länk.
6 Referenssystem*	Rumsligt referenssystem* Temporalt referenssystem Referenssystemets omfattning*	Rumsligt referenssystem bör anges med EPSG-kod. Även indirekta referenssystem förekommer (se ISO 19112).
7 Kvalitetskrav*		Beskrivning av kvalitetskrav på datamängden anges enligt ISO 19157.
7.1 Datakvalitet	Fullständighet * Logisk konsistens* Lägesnoggrannhet* Tematisk noggrannhet* Temporal noggrannhet * Kvalitetsområde* (omfattning)	Benämningarna om lägesnoggrannhet följer standarden 19131, som skiljer sig från benämningarna i detta dokument. (Se avsnitt 4.5 för ingående parametrar.) Beskriv det kvalitetsområde (geografiskt eller tematisk) som datakvalitetskraven gäller för.

		För produkter kan ytterligare krav anges för prestanda, kapacitet och tillgänglighet.
8 Metadata*		Beskriv de metadata som registreras.
9 Tillhandahållande*	Sammanfattning Omfattning*	Beskrivning hur dataprodukten tillhandahålls.
9.1 Leveransformat*	Beteckning ** Version Specifikation Filstruktur Språk ** Teckenuppsättning	Dataformat för leverans.
9.2 Leveransmedium*	Indelningsalternativ** Medium Volym Övrig information	Medium för leverans enligt metadatastandard till exempel dvd, ftp, online med flera.
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.

* Obligatorisk uppgift enligt ISO 19131

** 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 40:2012, 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 dataprodukten innehåll och struktur med applikationsschema (UML rekommenderas) och objekttypskatalog.
4	Beskriv dataprodukten 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 dataprodukten 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.

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.

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

IT-utvecklaren

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

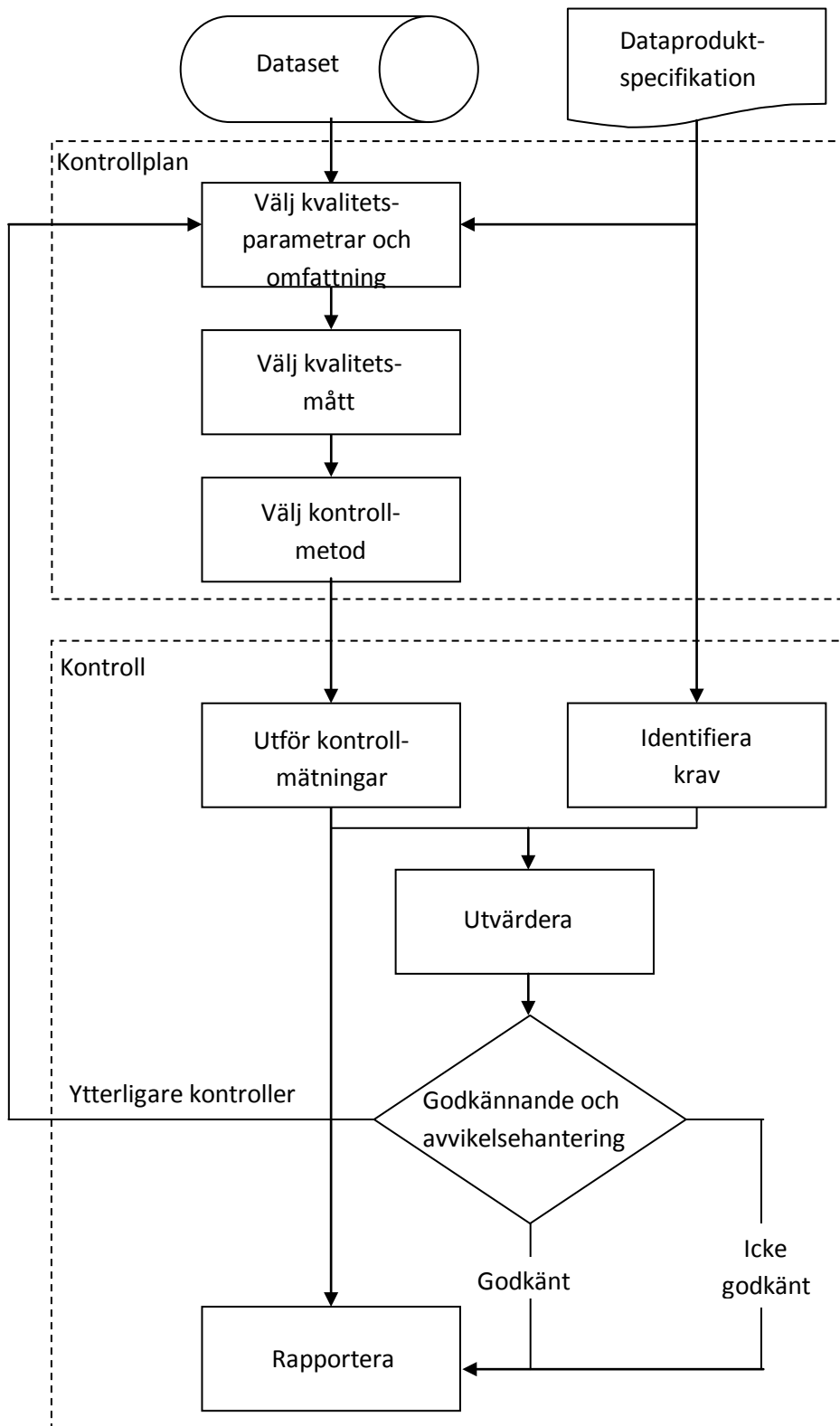
Läser applikationsscheman och information om tillhandahållande samt metadata.

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 nya norska standarden Geodatakvalitet)

En översikt av 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 kommande norska standarden "Geodatakvalitet".

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

Upprätta kontrollplan	
Välj kvalitetsparametrar och omfattning	Till exempel <i>lägesnoggrannhet</i> och <i>fullständighet</i> , se Tabell 4.5; 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 lägesnoggrannheten 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 kommande norska standarden "Geodatakvalitet".

Kvalitets-tema	Kvalitets-parameter	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	+	++	+		+				
Lägesnoggrannhet	Absolut	+			+	+	+	++		++
	Relativ	+		+				++		++
	Rasterdata	+			+	+		++		++
	(Tillförlitlighet)	+						++		++
Tematisk noggrannhet	Klassificering	+			+	++	+		++	
	Kvalitativa attribut	+				++			++	
	Kvantitativa attribut	+				++		+	++	++
Temporal noggrannhet	Tid	+	++	+						
	Temporal konsistens	+	++							
	Temporal validitet	+				++				
Användbarhet	Lämplighet	+								
	Aggregering	+	++							

++ Metoden är användbar
+ Metoden kan vara användbar

Relativ lägesnoggrannhet (lägesosäkerhet) ger ingen information om noggrannheten i referenssystemet. Kontroll av den absoluta lägesnoggrannheten är därför viktigare.

C Ett större tillämpningsexempel

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

C.1 Dataproduktspecifikationens krav

En tänkt dataproduktspecifikation har följande specificerade datakvalitet:

- **Felklassificerade objekt**, byggnader: $p_o = 0,03$ (3 %)
- **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 5.4.

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	Objektklassificering (kvalitativ)	856	80
Stödmurar	Fullständighet (kvalitativ)	58	13
Gränspunkter	Lägesosäkerhet och systematik (kvantitativ), grova fel (kvalitativ)	132	10 (20) 20

C.2 Test av objektklassificering

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

Tolerans, objektklassificering: 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, lägesosäkerhet och systematik

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

Avvikelser 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 kvalitén $p_o = 0,01$ (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.

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.

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

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)

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

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

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

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.

D Norskt register över kvalitetsmått

Denna bilaga är ett utdrag ur den kommande 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-standardens många illustrativa figurer som ett värdefullt komplement för att fullt ut förstå de olika måtten.

Bilagan kompletterar (och överlappar) avsnitt 4.5, och ger en samlad beskrivning av kvalitetsteman, -parametrar och -mått.

D.1 Inledning

Enligt datakvalitetsmodellen i ISO19157 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 Lägesnoggrannhet

Med *lägesnoggrannhet* menas hur bra ett objekts rumsliga läge är fastställt i förhållande till verkligheten (facit).

Kvalitetsparametrarna är *absolut lägesnoggrannhet*, *relativ lägesnoggrannhet* och *lägesnoggrannhet 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 noggrannhet

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

Klassificeringsnoggrannhet

Med *klassificeringsnoggrannhet* 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 noggrannhet, kvalitativa attribut

Den tematiska noggrannheten 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 noggrannhet, kvantitativa attribut

Den tematiska noggrannheten för kvantitativa attribut avser noggrannheten 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 noggrannhet

Temporal noggrannhet är datakvalitetsegenskaper som definierar tid eller tidsberoenden mellan objekt.

Tidsnoggrannhet

Med tidsnoggrannhet avses noggrannheten/osäkerheten i tidsmätningen jämfört med de sanna värdena.

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 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 attributnoggrannhet:

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