BeFo

STIFTELSEN BERGTEKNISK FORSKNING
ROCK ENGINEERING RESEARCH FOUNDATION

EFFEKTIV SIMULERING AV BROTTSANNOLIKHET:

Förstudie av AWH-metodens möjligheter inom bergbyggande

Jack Lidmar
Catrin Edelbro
Jessa Vatcher
Johan Spross

BeFo Rapport 240

STIFTELSEN BERGTEKNISK FORSKNING
ROCK ENGINEERING RESEARCH FOUNDATION

EFFEKTIV SIMULERING AV BROTTSANNOLIKHET:

Förstudie av AWH-metodens möjligheter inom bergbyggande

Efficient simulation of failure probability:

Prestudy of the AWH method's potential in rock engineering design

Jack Lidmar, KTH Royal Institute of Technology

Catrin Edelbro, Itasca Consultants AB

Jessa Vatcher, Itasca Consultants AB

Johan Spross, KTH Royal Institute of Technology

BeFo Rapport 240
Stockholm 2023
ISSN 1104-1773
ISRN BEFO-R-240-SE

FÖRORD

Denna förstudie behandlar potentialen i en ny metod för brottsannolikhetsberäkning, kallad *Accelerated Weight Histogram method*. Forskningen har utförts som ett samverkansprojekt mellan KTH Jord- och bergmekanik, KTH Fysik och Itasca. Forskningen har bedrivits inom det BeFo-finansierade projektet *Effektiva simuleringar av brottsannolikhet – etapp 1: förstudie*. Arbetet har pågått under 2020–2023.

Arbetet med förstudien har stöttats av John Leander (KTH) och Marie Westberg Wilde (AFRY / KTH), samt en referensgrupp, som författarna och BeFo riktar ett särskilt tack till. Följande personer medverkade i referensgruppen: Miriam Zetterlund (Tyréns), Alexandra Krounis (Sweco), Tobias Gasch (COMSOL), Eleni Gerolymatou (Chalmers), Håkan Stille (KTH), Lars Olsson (Geostatistik) och William Bjureland/Patrik Vidstrand (BeFo).

Stockholm,

Patrik Vidstrand

PREFACE

This pre-study investigates the potential of a new method for reliability-based design, called the Accelerated Weight Histogram method. The research was conducted as a collaboration between KTH Soil and Rock Mechanics, KTH Physics, and Itasca, within the project Efficient simulation of failure probabilities: Part 1: pre-study during 2020–2023.

The research was supported by John Leander (KTH) and Marie Westberg Wilde (AFRY / KTH), as well as a reference group. Their support is gratefully acknowledged. The reference group consisted of Miriam Zetterlund (Tyréns), Alexandra Krounis (Sweco), Tobias Gasch (COMSOL), Eleni Gerolymatou (Chalmers), Håkan Stille (KTH), Lars Olsson (Geostatistik) and William Bjureland/Patrik Vidstrand (BeFo).

Stockholm,

Patrik Vidstrand

SAMMANFATTNING

Vid dimensionering av undermarksanläggningar i berg behöver man hantera stora osäkerheter, eftersom man normalt inte har ekonomisk eller praktisk möjlighet att skaffa sig detaljerad kunskap om bergets egenskaper och spänningstillstånd. Ett sätt att stringent beakta dessa osäkerheter i dimensioneringen är att använda sig av sannolikhetsbaserade metoder, där osäkerheten modelleras som stokastiska variabler. Det möjliggör en beräkning av konstruktionens brottsannolikhet. Det finns dock i dagsläget ett behov av dels effektivare beräkningsmetoder för brottsannolikhet, dels en tydligare definition av brott i konstruktion hos en undermarksanläggning. Denna förstudie introducerar en ny Monte Carlo-baserad simuleringsmetod för beräkning av brottsannolikhet, på engelska benämnd *Accelerated Weight Histogram method* (AWH-metoden). Rapporten undersöker dess användbarhet på olika typer av gränstillstånd och ger dessutom ett mer detaljerat beräkningsexempel på dess tillämpning vid dimensionering av tunnelförstärkning. För detta beräkningsexempel introduceras en ny definition av brott i en tunnelkonstruktion, som är tänkt att avspegla tunnelkollaps i form av ett globalt bärighetsbrott. Bärighetsproblemet analyseras i en tunnelmodell i programvaran FLAC3D. Fördelen med att definiera brott som ett globalt bärighetsbrott är att den beräknade brottsannolikheten kan jämföras mot den tillåtna brottsannolikhet för konstruktioner som anges i standarder såsom Eurokoderna.

Resultatet av utförda analyser visar att AWH-metoden för de undersökta gränstillstånden presterar jämbördigt till bättre än delmängdssimulering (*subset simulation*), vilket är en relativt ny metod med liknande användningsområde. Avseende modelleringen av bergmassans beteende finns ett behov av att studera hur ingående parametrar kan beskrivas som sannolikhetsfördelningar. Förstudien visar att AWH-metoden kan bli ett användbart verktyg bland de sannolikhetsbaserade beräkningsmetoderna, men att det finns ett större antal detaljer som kan finslipas. Exempelvis är beräkningstiden fö närvarande ett praktiskt problem, vars möjliga lösning dock diskuteras i rapporten. Utöver det behövs även praktiska rekommendationer för användaren.

Nyckelord: Sannolikhetsbaserad dimensionering, tunnel, AWH-metoden, brottsannolikhet

SUMMARY

Design of underground facilities in rock implies management of large uncertainty, as there usually is not economically or practically feasible to collect detailed knowledge about the properties and stress conditions of the rock mass. One way to stringently consider these uncertainties in design is to apply reliability-based methods, in which the uncertainties are modelled as stochastic variables. This facilitates the calculation of the probability of structural failure. There is however currently a need for more efficient calculation or simulation methods to assess failure probabilities, as well as for a clearer definition of failure of a tunnel structure. This pre-study introduces a novel Monte-Carlo-based simulation method for assessment of failure probabilities, called the Accelerated Weight Histogram (AWH) method. Its applicability to a number of different types of limit states is investigated, including a tunnel design problem. In the latter case, a novel definition of structural failure of the tunnel was used, which describes failure as a global instability problem. The instability problem is analysed in the software FLAC3D. The advantage of this failure definition is that the calculated failure probability straightforwardly can be compared against established target failure probabilities stated in design codes such as the Eurocodes.

The result of the performed analyses shows that the AWH method for the analysed limit state performs equally well to better, compared to subset simulation, which is a relatively new method used for similar assessments. Concerning the modelling of the rock mass behaviour, there is a need to study how to describe the affecting parameters as probability density functions. The study indicates that the AWH method can become a useful tool among the reliability-based simulation methods, although there are still considerable amounts of details that can be improved upon in future studies. For example, the simulation time is a practical problem, the potential solution of which however is discussed in the report. In addition, there is a need to provide practical recommendations for the user.

Keywords: Reliability-based design, tunnel, AWH method, probability of failure

INNEHÅLL

1 INLEDNING

1.1 Problemställning

Bergbyggande karaktäriseras av att stora osäkerheter kring bergets egenskaper behöver hanteras när anläggningen ska dimensioneras. Osäkerheterna rör till stor del en brist på kunskap om bergets faktiska egenskaper, eftersom det i förväg ofta är svårt att utföra noggranna undersökningar långt in i bergmassan. Två tekniskt utmanande situationer som kan uppstå vid bergbyggande är dels ytligt belägna tunnlar med stor spännvidd (relativt bergtäckningen), dels passage av svagt berg på litet eller stort djup. Bägge problemen ställer krav på valet av analysmetod vid dimensionering. Vidare erfordras att man kan ge gränsen för acceptabelt beteende en relevant definition, exempelvis kollaps eller otillräcklig brukbarhet, samt att en tillräcklig säkerhetsmarginal mot detta gränstillstånd kan uppnås.

Traditionellt sett har man använt deterministiska säkerhetsfaktorer för att ge konstruktionen tillräckliga säkerhetsmarginaler. I en sådan analys beskrivs varje relevant bergmekanisk parameter endast med ett enskilt värde: ett medelvärde eller ett på annat sätt karaktäristiskt värde. Principiellt kan en sådan säkerhetsfaktor, *SF*, som baseras på medelvärden, för ett gränstillstånd beskrivas som:

$$SF = \frac{\mu_{\mathrm{R}}}{\mu_{\mathrm{S}}} \geq SF_{\mathrm{krav}}, \qquad (1)$$

där μ_{R} betecknar medelvärdet på den totala kapaciteten hos konstruktionen i analyserad brottmod, och μ_{S} medelvärdet på den totala lasten i brottmoden. Båda räknas fram från underliggande bergmekaniska parametrarnas medelvärden. Den beräknade *SF* ska vara minst lika stor som SF_{krav}, vars värde normalt fastställs av behörig myndighet eller i branschriktlinjer.

En sådan deterministisk analys utgår endast från medelvärden eller annat karaktäristiskt värde som aktuell dimensioneringsstandard angett (exempelvis en 5%-fraktil). Därmed saknas möjlighet att stringent beakta rådande nivå på osäkerheten för olika bergmekaniska parametrar. En deterministisk säkerhetsfaktor blir således ett trubbigt verktyg när man eftersträvar en verifiering av att konstruktionen har tillräckligt låg *sannolikhet* för kollaps, eller annat oönskat beteende såsom otillräcklig brukbarhet. Det beror på att två dimensioneringslösningar med olika stor osäkerhet kring bergets egenskaper ändå kan få samma beräknade säkerhetsfaktor om medelvärdena är desamma. Inom forskningen på dimensioneringsprinciper har man därför kommit att anse att sannolikheten för det

oönskade beteendet är ett bättre mått på hur riskfylld konstruktionen är, än den deterministiska säkerhetsfaktorn är. Det kan dock noteras att en deterministisk säkerhetsfaktor kan ge säkerhetsmarginal åt andra typer av osäkerheter såsom mänskliga fel (Doorn & Hansson, 2011). Forskningen om konstruktioners säkerhet har ändå under de senaste 50 åren gått mot sannolikhetsbaserade principer och utvecklat mer eller mindre avancerade och precisa sannolikhetsbaserade dimensioneringsmetoder. Att undersöka hur man kan dimensionera tunnlar för olika bergmekaniska problem med sannolikhetsbaserade metoder är därför intressant ur både vetenskaplig och praktisk synvinkel, särskilt för de två särskilt utmanande förutsättningarna som nämndes ovan: den vida, ytliga tunneln och den ytligt eller djupt liggande tunneln i svagt berg. Detta eftersom avancerade – noggranna – beräkningar är mest relevanta när man ställs inför de svåraste förutsättningarna.

Forskningen om sannolikhetsbaserad dimensionering började i mitten av 1900-talet. Bland tidiga pionjärer finns även svenska forskare (Kjellman & Wästlund, 1940). Forskningen kom sedan igång ordentligt på 1960- och 1970-talet och då inom stål- och betongbyggnad (se t.ex. Cornell, 1969). Några viktiga koncept som utvecklades var beskrivningen av ingenjörsproblemet som en så kallad gränsfunktion, $G(\mathbf{X})$, där oönskat beteende (”brott i konstruktionen”) definierades som att funktionen får utfallet $G(\mathbf{X}) \leq 0$. Vektorn $\mathbf{X} = [X_1, X_2, \ldots, X_m]$ innehåller här alla osäkra parametrar i beräkningen beskrivna som stokastiska variabler. Exempelvis kan bergmekaniska egenskaper beskrivas som normalfördelningar eller lognormalfördelningar. Den sannolikhetsbaserade motsvarigheten till den deterministiska verifieringen i ekv. (1) kan då uttryckas som att den beräknade brottsannolikheten p_{F} blir:

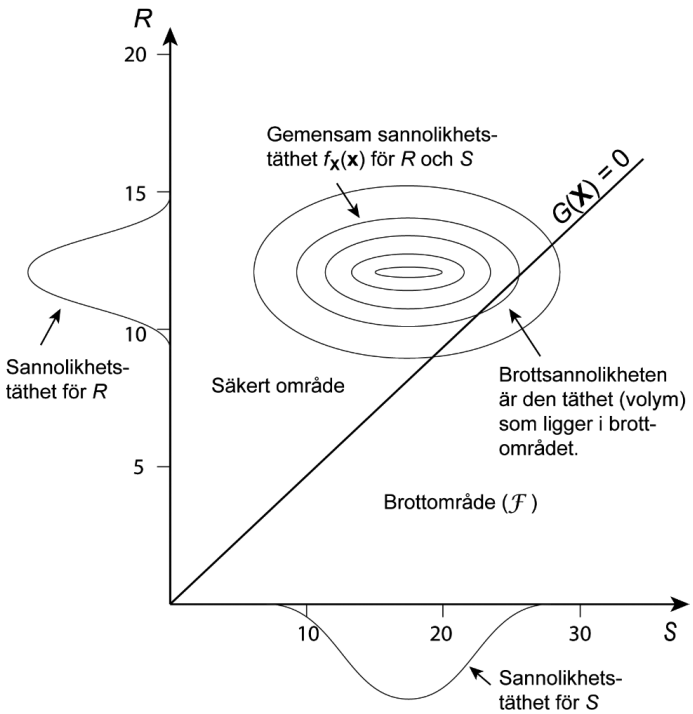
$$p_{\mathrm{F}} = \mathrm{P}[G(\mathbf{X}) \leq 0], \tag{2}$$

vilken ska vara mindre än eller lika med den tillåtna brottsannoliketen $p_{\mathrm{F,krav}}$. Enklast möjliga gränsfunktion är $G(\mathbf{X}) = R - S$, där R är en stokastisk variabel som beskriver mothållet och S är en stokastisk variabel som beskriver lasten. Ekv. (2) beskriver då sannolikheten att den osäkra lasten S är större än det osäkra mothållet R för konstruktionen, vilket illustreras i Figur 1.

I det generella fallet med många osäkra parametrar i vektorn $\mathbf{X} = [X_1, X_2, \ldots, X_m]$ kan ekv. (2) formuleras som en multidimensionell integral:

$$p_{\mathrm{F}} = P[G(\mathbf{X}) \leq 0] = \int \ldots \int_{G(\mathbf{X}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \tag{3}$$

där $f_{\mathbf{X}}(\mathbf{x})$ är den multivariata täthetsfunktionen för de beaktade stokastiska variablerna. Notera att integralens domän ges av brotthändelsen $\mathcal{F} = \{G(\mathbf{X}) \leq 0\}$, alltså det av gränsfunktionen avskurna området för brott i Figur 1.



Figur 1. *Definitionen av brottsannolikhet.*

En beräknad brottsannolikhet p_{F} jämförs sedan med en tillåten brottsannolikhet, $p_{\mathrm{F,T}}$, vilken normalt fastställs av behörig myndighet, exempelvis som en del i en utgiven dimensioneringsstandard. Eurokoderna är ett exempel på en sådan standard och där anges tillåtna brottsannolikheter på konstruktioner i storleksordningen 10^{-5} till 10^{-7}, beroende på vilka konsekvenser som uppkommer i händelse av brott (CEN, 2002).

För att åstadkomma en tillräckligt låg brottsannolikhet hos den dimensionerade konstruktionen har olika strategier utvecklats. De kan kategoriseras i:

- semi-probabilistiska metoder,

- analytiska metoder, och
- numeriska simuleringsmetoder.

Alla metoder har sina för- och nackdelar, vilka diskuteras kortfattat i det följande. För användaren gäller det att välja rätt metod till rätt problem, så att den beräknade brottsannolikheten blir tillräckligt noggrann utan att beräkningen tar för lång tid.

Till de semi-probabilistiska metoderna hör partialkoefficientmetoden, som finns i några olika varianter. Partialkoefficientmetoden används mycket inom betong- och stålbyggnad och fungerar där tämligen väl. Det är exempelvis den metod som rekommenderas som förstahandsalternativ i Eurokoderna (CEN, 2002) för dimensionering av byggnadsverk i allmänhet. Den används även för dimensionering av geotekniska konstruktioner i jord i Eurokod 7 (CEN, 2004), även om där finns en del teoretiska utmaningar framför allt för samverkanskonstruktioner. Den används dock inte inom bergbyggande. Den begränsade forskning som finns indikerar att nämnda teoretiska utmaningar finns även inom bergbyggande, och att de eventuellt är betydligt större där, se exempelvis Johansson et al. (2016).

De analytiska metoderna blir ofta approximativa, utom för vissa enkla gränsfunktioner. Om gränsfunktionen är tidskrävande att utvärdera, exempelvis för att den beror av en numerisk modell (t.ex. FEM), kan gränsfunktionen approximeras som en surrogatmodell (responsyta) med hjälp av en optimeringsalgoritm (Myers et al. 2009). Brottsannolikheten för den approximerade gränsfunktionen kan sedan lösas med en analytisk metod. Ett enkelt exempel ges av Damasceno et al. (2020), som visar hur en surrogatmodell kan användas tillsammans med den analytiska metoden FORM, för att beräkna brottsannolikheten för upplyftning av bergmassan ovanför ett trycksatt underjordiskt gaslager.

De numeriska simuleringsmetoderna är vanligtvis utvecklade från vanlig Monte Carlosimulering, i vilken man genererar ett stort antal slumptal från de stokastiska variablerna i $\mathbf{X}$ och undersöker i hur stor andel av fallen som gränsfunktionen tar värden $G(\mathbf{X}) \leq 0$. I vanlig Monte Carlo-simulering kräver detta normalt ett extremt stort antal utvärderingar av gränsfunktionen, eftersom en brottsannolikhet på säg 10^{-5} = 1 / 100 000 bokstavligen innebär att sannolikheten att uppfylla $G(\mathbf{X}) \leq 0$ i genomsnitt bara sker i en av hundratusen beräkningar. Därför har man utvecklat olika strategier för att effektivisera Monte Carlo-metoden, både inom forskning om konstruktioners säkerhet, men även inom andra forskningsfält där man behöver simulera sannolikheter. Ett exempel på en sådan mer avancerad Monte Carlo-metod är delmängdssimulering (eng: *subset simulation*), vars möjliga användning inom bergbyggnad visats i Damasceno et al. (2019) och Spross et al. (2022).

I denna förstudie har vi hämtat inspiration från statistisk fysik, där Lidmar (2012) utvecklade en ny Monte Carlo-metod, som han på engelska kallar Accelerated Weight Histogram method, fortsättningsvis benämnd AWH-metoden. Metoden koncentrerar beräkningarna adaptivt till de intressanta parameterområdena, vilket för tidigare undersökta tillämpningar i statistisk fysik och biofysik drastiskt minskat beräkningsbördan. Eftersom AWH-metoden utvecklades för att simulera sannolikheten för ovanliga händelser, finns en potential för att använda denna metod även för beräkningar av brottsannolikhet.

Johansson et al. (2016) ger en omfattande litteraturstudie av användningen av sannolikhetsbaserade dimensioneringsmetoder inom bergbyggande, så områdets historiska utveckling återupprepas inte här. En viktig förutsättning för tillförlitlig användning av sannolikhetsbaserade analyser inom bergbyggande är dock fortfarande olöst: hur ska brott i en bergmassa definieras, d.v.s. hur ska gränsfunktionen ställas upp? Ur ett konceptuellt perspektiv vore det önskvärt om gränsfunktionen beskrev ett tydligt observerbart brott-tillstånd – exempelvis tunnelkollaps – men detta beteende är inte helt enkelt att fånga och kvantifiera i en numerisk modell. Alternativa brottgränser har därför föreslagits och använts i tidigare forskning, exempelvis olika kriterier för en tillåten töjning eller deformation (se t.ex. Stille et al. 2005; Zhang & Goh, 2012; Bjureland et al. 2017). För dessa olika sätt att definiera brottgränsen blir det dock en påtaglig skillnad i hur allvarligt ett uppnått gränstillstånd är för konstruktionen, vilket gör att de olika brottgränsdefinitionerna rimligen ska associeras med olika tillåtna $p_{\mathrm{F,krav}}$. Att hitta tydliga och lättanvända definitioner av brottgränsen i en bergmassa är en förutsättning för att möjliggöra praktisk användning av sannolikhetsbaserad dimensionering inom bergbyggande, eftersom den tillåtna $p_{\mathrm{F,krav}}$ kan behöva ges olika värden beroende på hur brottgränsen definierats.

1.2 Studiens syfte och omfattning

Syftet med denna förstudie är att undersöka AWH-metodens funktionalitet för beräkningar av konstruktioners brottsannolikhet och jämföra den med andra befintliga metoder. Studien undersöker dels metodens funktionalitet på några principiellt olika gränsfunktioner ur ett teoretiskt perspektiv, dels hur metoden presterar i ett praktiskt bergmekaniskt beräkningsexempel, där vi beräknar brottsannolikheten för en tunnel vars beteende beskrivs av en numerisk modell i FLAC3D. För det praktiska exemplet visar vi även ett förslag på hur brott i en bergmassa kan beskrivas när man använder en numerisk modell.

1.3 Studiens utförande och rapportens upplägg

Studien började med att anpassa den generella algoritmen för AWH-metoden till att beräkna just brottsannolikheter. Därefter skedde ett förbättringsarbete på algoritmen, för att försöka minska antalet utvärderingar av gränsfunktionen men ändå bibehålla tillräcklig noggrannhet. Detta i syfte att hålla nere beräkningstiden, vilket är särskilt påtagligt när AWH-metoden kombineras med numeriska modeller. Detta beskrivs i kapitel 2.

Parallellt med utvecklingen av AWH-metoden utfördes även ett arbete för att på ett rättvisande sätt definiera brottbeteende i en bergmassa som beskrivs av en numerisk modell. Detta beskrivs i kapitel 3.

Kapitel 4 ger en sammanfattning av de två vetenskapliga artiklar som i detalj redogör för AWH-metoden och dess möjliga tillämpning på brottsannolikhetsberäkningar. Artikel A är en konferensartikel som jämför AWH-metoden med delmängdssimulering för brottsannolikhetsberäkningar med några teoretiska gränsfunktioner. Artikel B är en tidskriftsartikel där vi vidareutvecklar AWH-metoden för att kunna beräkna brottsannolikheten för tunnelstabilitet. Kapitel 5 diskuterar hur AWH-metoden presterade i de undersökta fallen, vilket följs av slutsatser och rekommendationer för fortsatt forskning i kapitel 6.

1.4 Avgränsningar

I denna förstudie har vi funnit det lämpligt att i första hand analysera det är det principiella beteendet för AWH-metoden. Därför har ett numeriskt ”enkelt” beräkningsexempel använts, där berget kan analyseras som ett kontinuum, utan explicit representation av sprickor. Tunnlar som passerar partier med berg av dålig kvalité anses representativa för det numeriska beräkningsexempel som valts ut. Förutsättningarna för att använda AWH-metoden tillsammans med mer komplexa numeriska modeller diskuteras dock i kapitel 5.

2 AWH-METODEN

2.1 Bakgrund

Intresset för att simulera komplexa probabilistiska fysikaliska eller statistiska modeller har vuxit enormt under de senaste decennierna, och en uppsjö av olika varianter på Monte Carlo metoder har utvecklats. I detta sammanhang utvecklades AWH-metoden (Lidmar, 2012) för att förbättra samplingen i system med komplicerade energilandskap, till exempel oordnade magnetiska material (spinglasmodeller) och för att räkna ut fria energier för olika tillstånd. Metoden har sedan dess framgångsrikt anpassats för biofysikaliska tillämpningar i kombination med atomistiska molekyldynamik-simuleringar, till exempel av proteiner och andra biopolymerer. Gemensamt för dessa problem är att det ofta finns många metastabila tillstånd separerade av höga energibarriärer som har låg sannolikhet att korsas. För att effektivt kunna sampla dessa sällsynta övergångar modifierar vi i AWH-metoden denna sannolikhet adaptivt under simuleringens gång så att de inträffar tillräckligt ofta.

Vid beräkning av konstruktioners brottsannolikhet är vi på samma sätt ute efter att simulera sällsynta händelser, och just beräkningen av brottssannolikheter har många likheter med fria energiberäkningar i statistisk fysik. I det senare fallet är de mikroskopiska tillstånden fördelade enligt en Boltzmannfördelning

$$\mathrm{P}(x) = \frac{\exp(-E(x)/k_B T)}{Z} \tag{4}$$

där $E(x)$ är energin, k_B är Boltzmanns konstant, T är temperaturen och Z är en normeringskonstant. Precis som i flera andra avancerade metoder för att beräkna brottsannolikhet bygger AWH-metodens Monte Carlo-simulering på så kallade Markovkedjor. För att simulera den statistiska modellen med Markovkedjor behöver man i regel inte känna till normeringskonstanten Z, men den är av stort fysikaliskt intresse då den är direkt relaterad till den så kallade fria energin[1] via $F = -k_B \mathrm{T} \ln Z$. Analogt med detta gäller att fördelningen av parametrar $\boldsymbol{x}$ som ger upphov till brott i ett mekanistiskt problem (alltså att de ligger i brottregionen $\mathcal{F}$ i Figur 1),

$$P(\boldsymbol{x}|\mathcal{F}) = \frac{P(\mathcal{F}|\boldsymbol{x}) f_X(\boldsymbol{x})}{P(\mathcal{F})} = \frac{\mathbb{1}(\boldsymbol{x} \in \mathcal{F}) f_X(\boldsymbol{x})}{P(\mathcal{F})} \tag{5}$$

[1] Helmholtz fria energi ges as $F = U - TS$ där U är (inre) energin, T temperaturen och S entropin. Skillnaden i fri energi mellan två jämviktstillstånd ger en undre gräns för hur stort arbete som krävs för att åstadkomma en transformation mellan dessa (eller en övre gräns för hur mycket som kan utvinnas).

har brottssannolikheten $P(\mathcal{F})$ som okänd normeringskonstant. $\mathbb{1}(\cdot)$ betecknar här indikatorfunktionen som är lika med 1 om argumentet är sant, annars 0. Även inom Bayesiansk statistik är man ofta intresserad av normeringskonstanterna eftersom de möjliggör validering och jämförelse av olika statistiska modeller. Ekvation (5) påminner i sin uppställning faktiskt om Bayes teorem, vilket gör att man möts av liknande svårigheter vid beräkning.

2.2 Kort om AWH-metoden

Normeringskonstanter är typiskt besvärliga att beräkna och de flesta simuleringsbaserade metoder (inklusive AWH) kan bara uppskatta kvoter mellan normeringskonstanter för olika parametrar. För att beräkna absoluta värden behöver vi därför ett exakt referensfall att använda för att ”brygga” över simuleringen till det verkliga fallet vi är intresserade av. I AWH görs detta genom att introducera en hel familj av fördelningar $\mathrm{P}(x|\lambda_\mathrm{k})$, för olika parametervärden $\lambda_0, \lambda_1, \dots, \lambda_M$, som interpolerar mellan referensfördelningen i ena ändan och målfördelningen i den andra. I fall då brott kan definieras via en kontinuerlig gränsfunktion $G(\boldsymbol{x})$, är det praktiskt att välja familjen

$$\mathrm{P}(\boldsymbol{x}|\lambda_k) = \frac{\mathbb{1}(G(\boldsymbol{x}) < \lambda_k) f_X(\boldsymbol{x})}{\mathrm{P}(\mathcal{F}_k)} \tag{6}$$

där vi nu istället för den verkliga brotthändelsen $\mathcal{F} = \{G(\boldsymbol{X}) < 0\}$ i ekv. (5) får beakta den händelsen $\mathcal{F}_k = \{G(\boldsymbol{X}) < \lambda_k\}$, som är troligare att inträffa eftersom λ_k tar värden mellan noll och oändligheten: $0 = \lambda_0 < \lambda_1 < \cdots < \lambda_M = \infty$. Vi får alltså en sekvens av mer och mer sannolika brott definierade av $\mathcal{F}_k$.

I andra fall, exempelvis om brottvillkoret bara säger om brott inträffar eller ej istället för att ge $G(\boldsymbol{x})$ ett explicit värde, kan det vara lämpligare att välja

$$P(\boldsymbol{x}|\lambda_k) = \frac{\mathbb{1}(G(\boldsymbol{x}) < 0) f_X(\boldsymbol{x}|\lambda_k)}{P(\mathcal{F}_k)} \tag{7}$$

där istället den underliggande fördelningen $f_X(\boldsymbol{x}|\lambda_k)$ anpassats så att brott blir mer eller mindre sannolikt beroende på λ_k.

I båda fallen ges den sökta brottsannolikheten av $P(\mathcal{F}_0)$, medan $P(\mathcal{F}_M) = 1$ är känd. AWH-metoden går sedan ut på att behandla λ_k som en extra stokastisk variabel och simulera en simultan sannolikhetsfördelning (eng: joint probability density function)

$$\mathrm{P}(\boldsymbol{x}, \lambda) \propto e^{f_k} \mathbb{1}(G(\boldsymbol{x}) < \lambda_k) f_X(\boldsymbol{x}) \tag{8}$$

där koefficienterna f_k väljs adaptivt under simulerings gång så att alla λ_k i genomsnitt besöks med en förutbestämd sannolikhet π_k. Simuleringen kommer alltså att hoppa fram och tillbaka mellan olika λ_k, dvs mellan olika nivåer av brottssannolikheter.

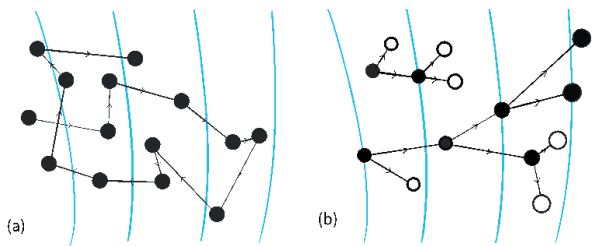
AWH-metoden är inte helt olik en annan vanlig metod för beräkning av brottssannolikheter, så kallad delmängdssimulering (eng: *subset simulation*) (Au & Beck, 2001), där man också simulerar en sekvens av nivåer. I fallet delmängdssimulering går man dock enbart i en riktning, från mer till mindre sannolika fall (Figur 2.). I AWH-metoden formas under simuleringens gång ett histogram W_k av vikter av besökta nivåer λ_k som används för att justera hyperparametrarna f_k tills histogrammet når målfördelningen, d.v.s. $W_k \approx N\pi_k$, se Figur 3 för ett exempel.

När detta är uppfyllt har simuleringen konvergerat och brottsannolikheten uppskattas då som

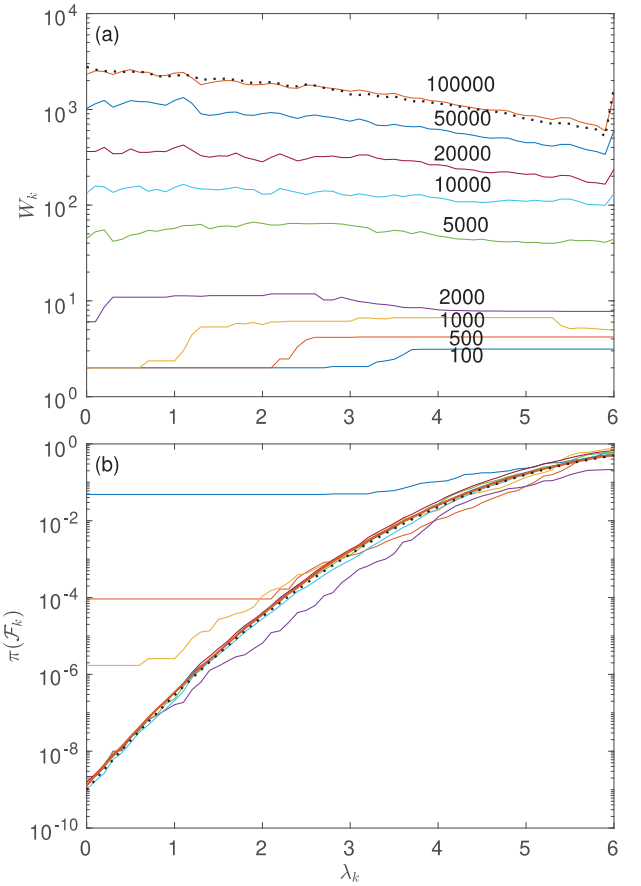
$$P(\mathcal{F}_k) \approx e^{(f_M - f_k)} \frac{\pi_M}{\pi_k} \tag{9}$$

På köpet får man brottsannolikheterna för alla mellanliggande tröskelnivåer λ_k, inte enbart för $\lambda_0 = 0$. Se Artikel A för ytterligare detaljer. I den artikeln har vi också gjort detaljerade jämförelser med delmängdssimulering för ett antal modellproblem, och funnit att AWH i flera fall producerar jämförbara eller bättre resultat.

Värt att notera är att man i AWH enkelt kan verifiera om simuleringen konvergerat som den ska genom att jämföra det empiriskt framräknade histogrammet W_k med målfördelningen π_k. För delmängdssimulering är det svårare att veta om resultaten är att lita på.



Figur 2. *Illustration av skillnaden mellan AWH (a) och delmängdssimulering (b). I AWH utförs en slumpvandring fram och tillbaka mellan olika nivåkurvor (i blått) av gränsfunktionen $\boldsymbol{G(x)}$, medan en delmängdssimulering startar med ett (stort) antal frön (sampel) som rör sig mot den alltmer osannolikare regionen där brottet sker.*



Figur 3. *(a) Viktningshistogram och (b) uppskattningar av brottssannolikheten som funktion av tröskelnivå* λ_k *för ett enkelt testfall* $G(\mathbf{X}) = 6 - (X_1 + X_2)/\sqrt{2}$*, där* X_i *är normalfördelade. Allteftersom simuleringen fortskrider byggs histogrammet upp och konvergerar slutligen mot en målfördelning som ges av den prickade linjen i (a). Samtidigt visas i (b) hur brottssannolikheten konvergerar mot det korrekta värdet* $P(\mathcal{F}) \approx 10^{-9}$ *vid* $\lambda = 0$.

3 BERGMEKANISK NUMERISK MODELLERING

För att kunna beräkna en brottsannolikhet för en tunnelkonstruktion behöver ett antal bergmekaniska aspekter definieras och implementeras i en numerisk modell. I detta kapitel diskuteras de bergmekaniska aspekterna på det beräkningsexempel som ingår i Artikel B. Detta eftersom fokus i Artikel B ligger på själva AWH-metodens förutsättningar att utföra beräkningen, snarare än bergmekaniska överväganden.

3.1 Förutsättningar och brottdefinition

Den numeriska modellen i en brottsannolikhetsberäkning används för att beskriva konstruktionens beteende. Tunnelmodellen som analyseras i Artikel B är "enkel" – i syfte att behålla fokus på AWH-metodens potential för beräkning av brottsannolikhet. Vi studerar en tunnelpassage i svagt berg på 300 meters djup. Bergförhållandena är inspirerade från passager i Nya Tunnelbanan och Förbifart Stockholm vilka temporärt krävt förstärkning med glasfiberbult och sprutbetong i stuff, långa spilingrör framför front, m.m., för att kunna passeras. ”Svagt berg” är i detta fall av sådan karaktär att det inte går att provtrycka och få värden genom standardmässiga laboratorietester, vilket naturligtvis ger stora osäkerheter i bestämningen.

Berget i Artikel B simuleras med en kontinuum-modell med idealplastisk materialmodell. Mohr-Coulombs skjuvbrottkriterium tillämpas med maximal hållfasthet för ingående parametervärden (kohesion, friktionsvinkel). Detta kriterium används eftersom det är det vanligast förekommande vid praktiskt dimensioneringsarbete och allmänt accepterat inom projektering av geokonstruktioner i Sverige. Dilationsvinkel och draghållfasthet simuleras som en deterministisk variabel för detta arbete.

Initiala huvudspänningar antas riktade horisontellt och vertikalt. Största horisontella primärspänning, σ_H antas vara orienterad vinkelrätt tunnelns axel och minsta horisontella primärspänning, σ_h, ortogonalt från det. Riktningen på huvudspänningar varieras inte. Vertikal initialspänning, σ_v, antas vara gravitativt beroende, men har trots det en relativt stor osäkerhet (såsom redovisas av t.ex. Palmström & Stille, 2015, samt underlaget till Tabell 1 nedan, som redovisas i Edelbro, m.fl., 2022). Initialspänningen i utförda beräkningar antas vara enligt framtaget primärspänningstillstånd för centrala delarna av Stockholm (Edelbro, m.fl., 2022) och med en magnitud enligt Tabell 1.

Tabell 1 *Underlag för åsättande av sannolikhetsfördelningar för spänningstillståndet. I Artikel B används log-normalfördelningar där medelvärdet motsvarar typvärdet i tabellen (Edelbro et al. 2022).*

Spänningssamband för centrala Stockholm*	σ_H [MPa]	σ_h [MPa]	σ_v
Låg	2.3 + 0.063 z	1.4 + 0.025 z	0.021 z
Typvärde (medelvärde)	3.6 + 0.080 z	3.2 + 0.035 z	ρg z
Hög	7.5 + 0.105 z	5.4 + 0.042 z	0.032 z

*Låg och hög indikerar realistiska skattningar på minsta och största värde på parametern.

Berget förstärks genom både bultning och ytförstärkning. Bultförstärkning simuleras i modellen med en idealplastisk materialmodell. Efter att sträckgräns uppnåtts tillåts sedan bulten töja sig (flyta) till dess att en specificerad töjningsgräns uppnås. Om töjningsgränsen överskrids går bulten av. Ytförstärkning med sprutbetong simuleras med elastisk materialmodell. Sprickor initieras i betongen vid uppnådd böjdrag- eller tryckhållfasthet.

Tunneln kan vara stabil trots att såväl berget som förstärkningen har genomgått plastiska deformationer. I arbetet studeras en nivå av brott med hänsyn till konsekvens. Brott som kan leda till utfall av berg och förstärkning anses ha hög och allvarlig konsekvens. Denna typ av brott antas ske när berget och förstärkningen deformerat tillräckligt mycket för att tappa den globala bärförmågan. Detta anser vi är en mer relevant beskrivning av brott i en tunnelkonstruktion, än exempelvis plasticering i en enskild punkt i bergmassan eller förstärkningen. Detta eftersom samhällets krav på säkerhet ofta uttrycks i termer av en (mycket liten) tillåten sannolikhet, $p_{F,T}$ för byggnadsverks kollaps. I denna rapport har vi valt att utvärdera den globala bärförmågan genom att studera om modellen går till jämvikt. I den programvara som använts för analysen bedöms instabilitet genom modellens kinetiska energi. Förändring i kinetisk energi mäts med konceptet *unbalanced force ratio* (Dawson et al. 1999). Om denna obalanserade kraft kommit ner till kravställd nivå är modellen i jämvikt; om den inte kommit ner till kravställd nivå, alternativt ökar, har modellen inte gått i jämvikt och brott registreras i beräkningen. Tillämpningen av denna princip för att definiera brotthändelsen $\mathcal{F} = \{G(\boldsymbol{X}) < 0\}$ i AWH-metoden diskuteras vidare i tillämpningsexemplet i Artikel B.

3.2 Modellering med FLAC3D

Den bergmekaniska spänningsanalysen utfördes i Itascas finita differens programvara *FLAC3D* (Itasca, 2019). En kvasi-3D modell skapades minimera körningstid. Kvasi-3D modeller är i sin grund 2D modeller med en liten tjocklek i den tredje dimensionen. I

denna typ av modeller används ett plant-töjnings- (eller plant-spännings-) symmetriförhållande för att representera ett tredimensionellt problem. Trots att en enskild körning tar kort tid så kräver sannolikhetsbaserade analyser ofta många iterationer. Därför var körtiden för en enskild modell en viktig faktor i besluten av vilken typ av modell och innehåll som skulle användas. Valet av *FLAC3D* som programvara baseras på dess integration med programmeringsspråket Python. Genom Python-programmering möjliggjordes automatiseringen av de sannolikhetsbaserade analyserna.

Programspråken Python och FISH (Itascas eget programspråk) integrerades och användes för att bygga, köra och utvärdera modellerna med indata slumpad från sannolikhetsfördelningar. I detta fall användes lognormalfördelningar på bergmassans indata och initialspänningstillståndet (se detaljer i Artikel B).Stegen i *FLAC3D*-simuleringarna är att AWH-metoden, som programmeras i Python, anger för vilka specifika indata som *FLAC3D*-modellen ska köras. Med hjälp av Python-programmering skickas den informationen in i en specifik modell i *FLAC3D*. Resultaten från *FLAC3D*-körningen rapporteras sedan tillbaka till Python-skriptet, där körningens resultat tas emot och behandlas i AWH-metodens algoritm, som sedan föreskriver nya indata för en ny körning i *FLAC3D*.

4 SAMMANFATTNING AV BIFOGADE ARTIKLAR

Två vetenskapliga artiklar har skrivits baserat på projektresultaten. De sammanfattas här. Artikel A finns av copyright-skäl enbart tillgänglig digitalt via länk, medan Artikel B har bilagts rapporten i sin helhet.

4.1 Artikel A

> Lidmar, J. Spross, J. & Leander, J. 2022 Accelerated Weight Histogram method for rare event simulations. *Proceedings of the 13th International Conference on Structural Safety & Reliability (ICOSSAR 2021-2022), 13-17 September 2022, Shanghai, China*, https://doi.org/10.48550/arXiv.2210.14537.

Artikeln beskriver hur den ursprungliga AWH-metoden kan omformuleras för att beräkna brottsannolikheter. Den nya metoden används sedan i två teoretiska beräkningsexempel för att undersöka metodens funktionalitet. AWH-metoden jämförs här med en konkurrerande Monte Carlo-baserad metod, så kallad delmängdssimulering (eng: *subset simulation*). Det första exemplet är av enklare natur. Det undersöker brottsannolikheten i en gränsfunktion byggd av en summa av normalfördelade stokastiska variabler. Det andra exemplet undersöker en gränsfunktion som representerar brottet i ett fiberknippe utsatt för drag, där varje tråd i knippet har samma styvhet men bryts vid en slumpmässig töjning. När lasten sakta ökar kommer de svagare trådarna gå av en efter en och lasten omfördelas till de kvarvarande. Sådana modeller utvecklades från början för att beskriva hur en bunt av trådar går sönder, men har fått spridning till många andra discipliner, även jordmekanik. Gränsfunktionen för ett fiberknippe är beräkningsmässigt mycket utmanande att beräkna brottsannolikheten för.

Resultatet av undersökningarna visar att för den enklare gränsfunktionen presterar AWH-metoden ungefär lika bra som delmängdssimulering, med ett litet övertag för delmängdssimuleringen. För det mer utmanande fiberknippeproblemet fick delmängdssimuleringen svårigheter och i detta fall presterade AWH-metoden betydligt bättre än konkurrenten.

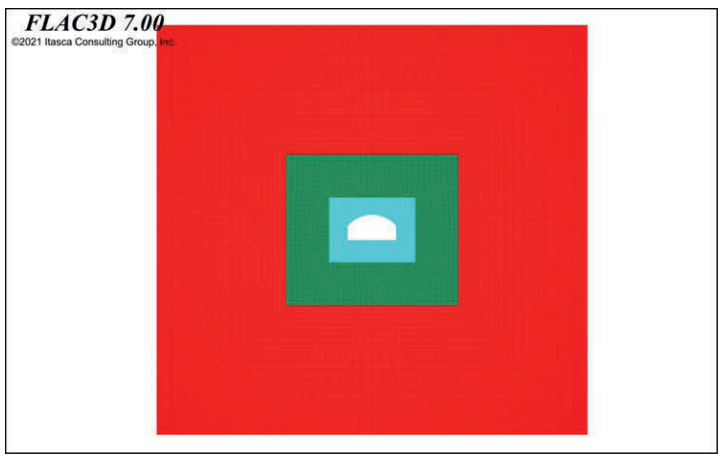
4.2 Artikel B

> Lidmar, J., Vatcher, J., Edelbro, C. & Spross, J. 2023. Estimation of small failure probabilities using the Accelerated Weight Histogram method. *Probabilistic Engineering Mechanics, 74*, 103501. https://doi.org/10.1016/j.probengmech.2023.103501

I den här artikeln beskrivs mer konkret hur AWH-metoden kan appliceras på mer komplicerade problem och i synnerhet för beräkning av sannolikhet för kollaps av ett tunneltvärsnitt modellerat med hjälp av FLAC3D. Ytterligare detaljer kring de bergmekaniska aspekterna ges i Artikel B, särskilt avseende parametrarnas sannolikhetsfördelningar.

Modellgeometrierna som användes i simuleringarna var typiska tunnelgeometrier (bredd 10 m) enligt Figur 4. Modelltjockleken ut ur planet var 1 m för att återskapa nästan fyrkantiga zoner som är optimalt för spänningsberäkningar. En zon i FLAC3D är en stängd geometrisk domän med noder hörnen och på ytorna som kopplar samman näraliggande zoner. Zonstorleken närmast tunneln hade sidlängden 1 meter och zonstorlekarna ökade sedan gradvis längre bort från tunnelranden för att minska körtiden. Rullstöd användes på alla sidor av modellen. Modellen kördes med möjlighet till små töjningar (positionerna i nätet uppdateras inte fysiskt) för att kunna jämföra AWH metodens resultat med programvarans inbyggda metod för säkerhetsfaktor, se FLAC3D:s manual (Itasca, 2019).

En komplikation som diskuterats i kapitel 3 är att brottkriteriet inte ges av en kontinuerlig gränsfunktion, utan istället av en diskontinuerlig funktion. Det innebär för tunnelexemplet att man utifrån observerad *unbalanced force ratio* direkt anger i ekv. (5) värdet $\mathbb{1}(\boldsymbol{x} \in \mathcal{F}) = 1$ om jämvikt ej uppnåtts i körningen för körningens indata $\boldsymbol{x}$, annars 0 om jämvikt uppnåtts. Man kan alltså hoppa över det annars vanliga steget att utvärdera ett explicit numeriskt värde på gränsfunktionen $G(\boldsymbol{x})$ och jämföra om det är större eller mindre än gränstillståndet $G(\boldsymbol{x}) = 0$.



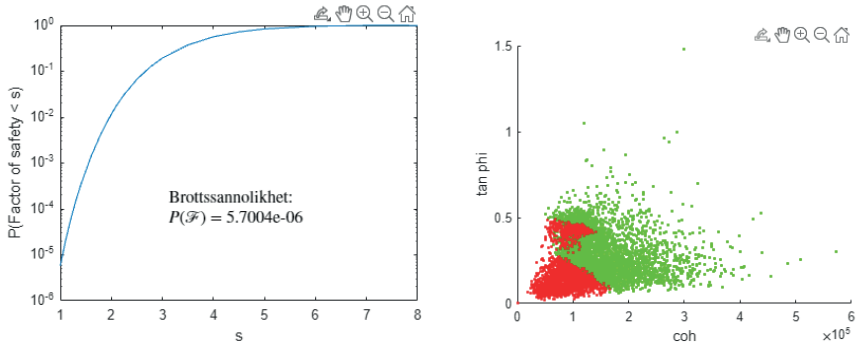
Figur 4. *Modellgeometri för det analyserade tunneltvärsnittet i FLAC3D.*

I våra *FLAC3D*-beräkningar definierade vi brott som att *unbalanced force ratio* ej understiger 10^{-5} under ett visst antal tidssteg. För att hantera detta i AWH-algoritmen introducerar vi en omskalningsparameter $s = e^{\lambda}$, som fungerar analogt med en säkerhetsfaktor i bergmekanisk dimensionering. Denna används för att skala om två av de viktigaste bergmekaniska parametrarna i problemet, nämligen kohesionen, c, och friktionskoefficienten, tan(φ), så att brott blir mycket sannolikare för stora s. På så vis fås en familj av sannolikhetsfördelningar $f_X(x|s_k)$ som interpolerar mellan det verkliga fallet vid $s_0 = 1$ ($\lambda = 0$) och allt sannolikare fall för $s_k > 1$. Denna familj av fördelningar kan användas med AWH-metoden så som beskrivet i kapitel 2.2. Totalt ingår 13 parametrar i den bergmekaniska numeriska modellen för tunneltvärsnittet, varav 10 är stokastiska och antas följa lognormala fördelningar med givna medelvärden och varianskoefficienter, se Tabell 2 för ett exempel. För ett givet värde på säkerhetsfaktorn s_k sker en omskalning enligt $\mathrm{c} = \mathrm{c}_{\mathrm{org}}/s_k$ och $\tan\varphi = \tan\varphi_{\mathrm{org}}/s_k$.

Resultatet av AWH-simuleringen ger då sannolikheten $P(\mathcal{F}|s_k)$ för brott som funktion av s_k, vilket kan tolkas som sannolikheten för att säkerhetsfaktorn skulle ha ett värde mindre än s_k. Exempel på hur det kan se ut visas i Figur 5.

Tabell 2. *Ingående parametrar i tunnelmodellen, med exempel på medelvärden och variationskoefficienter från ett beräkningsfall. Samtliga parametrar antas följa log-normala fördelningar.*

Bergmassans parametrar	**Medelvärde**	**Variations-koefficient**
Densitet (kg/m^3)	2650	0
Deformationsmodul (Pa)	$2 \cdot 10^9$	0.25
Tvärkontraktionstal	0.25	0.1
Kohesion (Pa)	$5 \cdot 10^5$	0.4
Friktionsvinkel (°)	33	0.4
Draghållfasthet (Pa)	200	0
Dilatationsvinkel (°)	2	0
Största horisontalspänning, σ_H (Pa)	$2.513 \cdot 10^7$	0.5
Gradient för ökning av σ_H med djup (Pa/m)	82 000	0.25
Minsta horisontalspänning, σ_h (Pa)	$1.13 \cdot 10^7$	0.5
Gradient för ökning av σ_h med djup (Pa/m)	32 800	0.25
Vertikalspänning σ_v (Pa)	$6.5 \cdot 10^6$	0.5
Gradient för ökning av σ_v med djup (Pa/m)	26 100	0.25



Figur 5. *Till vänster: Sannolikhet för säkerhetsfaktor ≤ **s** från AWH-simuleringar av tunnelmodell med parametrar givna i Tabell 2. Sannolikheten för brott är samma sak som sannolikheten att säkerhetsfaktorn är mindre än 1. Till höger: Bild av gränstillståndsytan projicerat på planet c–*tan*(φ). Röda punkter går till brott, gröna inte. Den oregelbundna formen på brottområdet indikerar att bergmodellens beteende är komplext när man ligger nära brottgränsen.*

I det aktuella problemet har kohesionen och friktionsvinkeln störst inverkan på brottssannolikheten och vi kan få en bild av gränstillståndsytan om vi projicerar de simulerade datapunkterna i det planet. Se högra grafen i Figur 5. Man ser här att gränsen mot brottsområdet (röda punkter) inte är helt skarp och att området inte heller är helt sammanhängande vilket gör problemet utmanande. Trots detta har simuleringen konvergerat mot en brottssannolikhet på ca $5.7 \cdot 10^{-6}$ efter 14 000 körningar i FLAC3D, vilket ses där vänstra figurens kurva möter vertikalaxeln ($s = 1$). Redan efter runt 2000 körningar fås en grov skattning som sedan förbättras efter hand. Tiden för att genomföra 14 000 körningar är i storleksordningen 100 timmar (ca 0,5 minut per körning).

Slutsatsen av simuleringarna i Artikel B är att AWH-metoden kan användas för att simulera brottsannolikheter i dimensionering av ett tunneltvärsnitt.

5 DISKUSSION

5.1 AWH-metodens funktionalitet

En aspekt som gör brottssannolikhetsberäkningar speciella är att det typiskt är mycket tidskrävande att utvärdera gränsfunktionen $G(\boldsymbol{x})$. Ofta krävs omfattande numeriska beräkningar för varje värde på indata $\boldsymbol{x}$, exempelvis som i den aktuella tunnelstudien där vi utförde FLAC3D beräkningar. Detta ställer stora krav på effektivitet hos simuleringsalgoritmen så att onödiga utvärderingar av $G(\boldsymbol{x})$ undviks. Beroende på problemets komplexitet kommer olika metoder att vara bäst lämpade. Om brottssannolikheten inte är alltför liten ($\gtrsim 0.05$) är det svårt att slå vanlig (”crude”) Monte Carlo-simulering, eftersom variablerna $\boldsymbol{x}$ då slumpas fram oberoende av varandra. Delmängdssimulering har på liknande sätt fördelen att den utgår från ett relativt stort antal oberoende ”frön” (sampel), som sedan flyttas mot brottregionen $G(\boldsymbol{x}) < 0$ med Markovkedje-Monte Carlo. Korrelationer mellan fröna byggs dock successivt upp för varje generation under simuleringens gång vilket minskar den fördelen för mer komplicerade problem. AWH arbetar mer sekventiellt med korrelerade sampel från en Markovkjedja, men vandrar fram och tillbaka över parameterområdet flera gånger vilket verkar leda till ett effektivare utforskande av parameterrummet för komplexa problem.

Oavsett vilken metod som används för brottssannolikhetsberäkning kan man inte vänta sig bättre resultat än vad den statistiska modellen kan beskriva. Speciellt bör man vara vaksam på att för problem där brottssannolikheten styrs av värdet på några få ingående parametrar finns en risk att resultatet blir väldigt känsligt för beteendet hos svansarna på fördelningarna av dessa parametrar. Det är då av yttersta vikt att dessa är tillförlitliga. Då det ofta finns betydande epistemiska osäkerheter gällande parametrarnas fördelning kan det vara lämpligt att använda hierarkiska probabilistiska modeller för att bygga in denna osäkerhet, alltså att sannolikhetsfördelningens varians i sig modelleras med en sannolikhetsfördelning. Användningen av detta tillvägagångssätt har diskuterats för bergparameterar av Bozorgzadeh et al. (2019).

I fall brottsannolikheten istället styrs av ett stort antal samverkande variabler har man inte denna känslighet, utan kan förvänta sig robustare resultat. Korrelation mellan variabler kan dock inverka på känsligheten. Ur denna synpunkt skulle det kunna vara intressant att studera mer detaljerade modeller för bergmassan som tar hänsyn till rumslig heterogenitet och korrelation.

Erfarenheten från beräkningarna i Artikel B visar att möjligheten att undersöka brottbeteendet binärt (d.v.s. med enbart ”brott” (1) eller ”ej brott” (0) som resultat i indikatorfunktionen $\mathbb{1}(\boldsymbol{x} \in \mathcal{F})$ i ekv. (5) istället för ett explicit beräknat värde på $G(\mathbf{X})$), gör AWH-metoden mycket intressanta för vissa gränstillstånd i en tunnel. Här har AWH-metoden en klar fördel gentemot exempelvis delmängdssimuleringar, men även metoder med surrogatmodeller, som båda kräver en kontinuerlig gränsfunktion. Ett exempel på en metod där surrogatmodeller används är *Adaptive directional importance sampling* (ADIS) (Grooteman, 2011), vars användbarhet för dimensionering av bergkonstruktioner visas i Damascenos (2022) doktorsavhandling. Med ADIS krävs mycket få simuleringar av den numeriska modellen, men den numeriska modellen behöver alltså kombineras med en kontinuerlig gränsfunktion, d.v.s. kunna ge ett explicit värde på $G(\boldsymbol{x})$, för att en surrogatmodell ska kunna skapas. Ett möjligt alternativ till AWH för att hantera binära brottvillkor (diskontinuerlig gränsfunktion) är så kallade sekventiella Monte Carlo metoder (SMC), se exempelvis Papaioannou et al., 2016, som i övrigt har många likheter med delmängdsimuleringar.

Det går inte att komma runt att många tusentals evalueringar av gränsfunktionen $G(\boldsymbol{x})$ ofta kommer att krävas för simuleringarna. Även för relativt förenklade numeriska modeller, som tunneltvärsnittet i Artikel B där varje enskild FLAC3D-beräkning tog runt 30 sekunder, får man att 3000 sampel tar ca ett dygn. För dessa simuleringar krävs ca 2000 sampel för en grov uppskattning av brottssannolikheten, men med fler får man bättre noggrannhet. Det är alltså fråga om relativt storskaliga datorberäkningar. Samtidigt är det inte nödvändigtvis så att själva gränsytan $G = 0$ har en extremt komplicerad form, utan i princip skulle den kunna beskrivas av en betydligt enklare surrogatmodell. En framkomlig väg, som kan undersökas i fortsatta studier, kan vara att först konstruera en surrogatmodell som ger en approximativ beskrivning av brottområdet $\mathcal{F}^S = \{G^S(\boldsymbol{X}) \leq 0\}$, och som är snabb att beräkna, och sedan använda den tillsammans med AWH för att beräkna brottssannolikheten. Här skulle det vara av intresse att studera möjligheten att i AWH-metoden kombinera en surrogatmodell och den ursprungliga numeriska modellen, så att antalet evalueringar av den senare minimeras.

5.2 AWH-metodens roll i tunnelingenjörens verktygslåda

Trenden inom utvecklingen av metoder för att verifiera konstruktioners säkerhet har inom såväl forskning som praktik rört sig bort från deterministiska säkerhetsfaktorer, mot en mer stringent hantering av osäkerhet och risk genom att istället sätta krav på tillåten brottsannolikhet. För konstruktioner av stål och betong och vissa geokonstruktioner i jord kan den semi-probabilistiska partialkoefficientmetoden utgöra ett tillräckligt bra verktyg för att verifiera att brottsannolikheten hos den projekterade konstrukt-

ionen är tillräckligt låg. Inom bergbyggande är det dock inte alls självklart att partialkoefficienter skulle fungera lika bra, vilket exempelvis BeFo-rapporten av Johansson et al. (2016) indikerar.

Om tunnelbyggnadsbranschen ska ta sig vidare från dagens deterministiska säkerhetsfaktorer och följa resten av byggbranschens övergång mot probabilistiska metoder, så kommer bra alternativ till partialkoefficientmetoden att behövas. Enligt vår bedömning finns det i dagsläget ingen universalmetod som kan beräkna brottsannolikheter enkelt och bra till alla olika typer av gränstillstånd som man har att hantera i ett tunnelprojekt i berg. Partialkoefficienter kan eventuellt visa sig fungera för verifiering av vissa gränstillstånd, men särskilt för samverkanskonstruktioner kommer det troligen behövas en fullt probabilistisk metod för verifiering.

Morgondagens tunnelingenjör kommer därför förmodligen att behöva förstå och behärska även några mer avancerade metoder, samt – inte minst – kunna avgöra vilken metod som passar bäst till vilken problemformulering. I valet av beräkningsmetodik kan man exempelvis behöva beakta osäkerheten i tillgängliga data, då åsättandet av fördelningar har stor påverkan på utfallet i denna typ av beräkningar. Såsom nämndes i rapportens inledning har alla befintliga metoder för beräkning av brottsannolikhet sina fördelar och nackdelar, avseende exempelvis noggrannhet, tidsåtgång och komplexitet. Vi ser här AWH-metoden som en av flera möjliga metoder som kan passa när en numerisk (Monte Carlo-baserad) simuleringsmetod är lämplig för att beräkna brottsannolikheten. I Artikel A såg vi exempelvis hur fiberknippeproblemet hanterades bättre av AWH-metoden än konkurrenten delmängdssimulering.

Vi tror dock att AWH-metoden och liknande metoder gör sig bäst om de kan integreras direkt i programvaran som utför beräkningen av tunnelmodellen, eftersom detta skulle förenkla användningen avsevärt. Den omfattande handpåläggning som idag krävs med kodning i Python är acceptabel vid utveckling av metoden i forskningsprojekt som detta, med för att skapa rimlig praktisk användbarhet anser vi att det är det rimligt att programvaruutvecklare integrerar sannolikhetsbaserade metoder i nya utgåvor av programvaror för numerisk modellering av konstruktioner.

6 SLUTSATS OCH REKOMMENDATION FÖR FORTSATT FORSKNING

Denna förstudie har undersökt AWH-metodens funktionalitet för att beräkna brottsannolikheter. Den bifogade Artikel A visade hur AWH-metoden kan formuleras för att beräkna en brottsannolikhet och jämförde metodens funktionalitet med delmängdssimulering för två olika gränsfunktioner. AWH-metoden gav jämbördiga till bättre resultat än konkurrenten. I Artikel B visar vi hur AWH-metoden kan användas för att beräkna brottsannolikheten för ett tunneltvärsnitt, där brottgränsen definierades på ett för forskningslitteraturen nytt sätt med hjälp av principen *unbalanced force ratio*. Detta möjliggjorde för oss att simulera ett globalt bärighetsbrott som ger ett representativt beteende för de små tillåtna brottsannolikheter ($\sim 10^{-5}$–10^{-7}) som anges i dimensioneringsstandarder som Eurokoderna. Detta gav en diskontinuerlig gränsfunktion. Trots denna utmanande beskrivning av brottgränsen kunde AWH-metoden modifieras för att utföra brottsannolikhetsberäkningen.

Avseende definitionen av brott med hjälp av *unbalanced force ratio* är detta ett lovande tillvägagångssätt för brottsannolikhetsberäkningar av tunnelkonstruktioner i allmänhet. Det ska alltså inte ses som en metodik som enbart är användbar tillsammans med AWH-metoden. Det finns här skäl att separat undersöka hur konceptet i detalj ska förstås och modelleras ur ett rent bergmekaniskt perspektiv. En tydlig och användbar definition av brott i bergmassa har nämligen efterfrågats länge bland forskare på dimensioneringsprinciper för bergbyggande, och att en sådan saknas utgör ett fundamentalt problem vid framtagande av dimensioneringsstandarder och kalibrering av säkerhetsnivåer. En näraliggande viktig frågeställning rör hur de bergmekaniska indataparametrarna lämpligast beskrivs som sannolikhetsfördelningar. Vid vilka förhållanden är det exempelvis viktigt att beskriva bergets heterogenitet med en rumslig variation hos en parameter och när kan en enklare homogen statistisk modell användas? När kan undersökningsdata lämpligen kombineras med erfarenhet, exempelvis med bayesianska statistiska metoder?

Tidsåtgången vid beräkningarna är i dagsläget en nackdel. AWH-metoden gör sig därför i dagsläget förmodligen bäst när det inte är så numeriskt krävande att utvärdera gränsfunktionen. Sådana problem kan ändå vara mycket svåra att simulera effektivt med andra metoder, om gränsfunktionen är komplex (såsom fiberknippemodellen i Artikel A) eller är en systemformulering av flera brottmoder. För numeriskt tyngre modeller bör man överväga att vidareutveckla AWH-metoden så att den kan hantera surrogatmodeller för att minska körtiden. En intressant möjlighet är att kombinera den ursprungliga modellen och surrogatmodellen i en och samma AWH-simulering. Detta

tillvägagångssätt skulle också kunna ge en betydande variansreduktion genom en omskrivning av brottsannolikheten som $P(\mathcal{F}) = P(\mathcal{F}^S) + \mathbb{E}[\mathbb{1}(x \in \mathcal{F}) - \mathbb{1}(x \in \mathcal{F}^S)]$. De båda termerna i väntevärdet kommer att vara korrelerade vilket leder till minskad varians vid uppskattning med Monte Carlo.

Sammanfattningsvis visar förstudien att AWH-metoden har en plats i verktygslådan bland de sannolikhetsbaserade dimensioneringsmetoderna. Det finns dock ett större antal detaljer som skulle kunna finslipas, för att få ner beräkningstiderna. Därefter bör man ta fram praktiska rekommendationer för användaren, exempelvis hur histogrammen ska tolkas och metodens styrparametrar åsättas och de bergmekaniska indataparametrarna beskrivas som sannolikhetsfördelningar.

7 REFERENSER

Au, S.-K., & Beck, J. L. 2001. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics, 16*(4), 263–277.

Bjureland, W., Spross, J., Johansson, F., Prästings, A., & Larsson, S. (2017). Reliability aspects of rock tunnel design with the observational method. *International Journal of Rock Mechanics and Mining Sciences, 98*, 102-110.

Bozorgzadeh, N., Harrison, J. P., & Escobar, M. D. 2019. Hierarchical Bayesian modelling of geotechnical data: application to rock strength. *Géotechnique, 69*(12), 1056-1070.

CEN. 2002. *EN 1990:2002 – Basis of structural design*. Brussels: European Committee for Standardisation.

CEN 2004. *EN 1997-1:2004 Eurocode 7: Geotechnical design – Part 1: General rules*. Brussels: European Committee for Standardisation.

Cornell, C.A. 1969. A probability-based structural code. *ACI Journal Proceedings, 66*(12), 974-984.

Damasceno, D. 2022. Modeling aspects of reliability-based design of lined rock caverns. Doktorsavhandling, TRITA-ABE-DLT-2242, KTH, Stockholm.

Damasceno, D. R., Spross, J., Johansson, F., and Johansson, J., 2019. Efficiency of subset simulation in the design of lined rock caverns for storage of hydrogen gas. *Proceedings of the 13th International Conference on Applications of Statistics and Probability in Civil Engineering*, ID 124, Seoul, South Korea.

Damasceno, D. R., Spross, J., and Johansson, F., 2020. Reliability-based design methodology for lined rock cavern depth using the response surface method. *Proceedings of the ISRM International Symposium – EUROCK 2020*, ID 163, Trondheim, Norway.

Dawson, E. M., Roth, W. H., & Drescher, A. 1999. Slope stability analysis by strength reduction. *Géotechnique, 49*(6), 835-840.

Doorn, N. & Hansson, S.O. 2011. Should probabilistic design replace safety factors? *Philosophy & Technology, 24*(2), 151-168.

Edelbro, C., Perman, F., Figueiredo, B. & Sjöberg, J. 2022. *Utvärdering och tolkning av initiala bergspänningar för Stockholm och Göteborg*. Rapport 231. BeFo, Stockholm.

Grooteman, F. 2011. An adaptive directional importance sampling method for structural reliability. *Probabilistic Engineering Mechanics, 26*(2), 134-141.

Itasca. 2019. *FLAC3D — Fast Lagrangian Analysis of Continua in Three-Dimensions*, Ver. 7.0. Minneapolis: Itasca Consulting Group, Inc.

Johansson, F., Bjureland, W. & Spross, J. 2016. *Application of reliability-based design methods to underground excavation in rock*, BeFo-rapport 155, BeFo, Stockholm.

Kjellman, W. & Wästlund, G. 1940. *Säkerhetsproblemet i byggnadskonsten*. Ingeniörsvetenskapsakademiens handlingar nr. 156. Generalstabens litografiska anstalts förlag, Stockholm.

Lidmar, J. 2012. Improving the efficiency of extended ensemble simulations: The accelerated weight histogram method. *Physical Review E, 85*(5), 056708.

Myers, R. H., Montgomery, D. C. & Anderson-Cook, C. M. 2009. *Response surface methodology: process and product optimization using designed experiments*. Hoboken, Wiley.

Palmström, A. & Stille, H. 2015. *Rock Engineering*. Thomas Telford, London.

Papaioannou, I., Papadimitriou, C., & Straub, D. 2016. Sequential importance sampling for structural reliability analysis. *Structural Safety*, *62*, 66–75. https://doi.org/10.1016/j.strusafe.2016.06.002

Spross, J., Gasch, T. & Johansson, F. 2022 Implementation of reliability-based thresholds to excavation of shotcrete-supported rock tunnels, *Georisk*, in press. DOI: 10.1080/17499518.2022.2046789

Stille, H., Holmberg, M., Olsson, L. & Andersson, J. 2005. *Dimensionering av samverkanskonstruktioner i berg med sannolikhetsbaserade metoder*. SveBeFo-rapport 70. BeFo, Stockholm.

Zhang, W., & Goh, A. T. 2012. Reliability assessment on ultimate and serviceability limit states and determination of critical factor of safety for underground rock caverns. *Tunnelling and Underground Space Technology, 32*, 221-230

Artikel A

Lidmar, J. Spross, J. & Leander, J. 2022 Accelerated Weight Histogram method for rare event simulations. *Proceedings of the 13th International Conference on Structural Safety & Reliability (ICOSSAR 2021-2022), 13-17 September 2022, Shanghai, China*, https://doi.org/10.48550/arXiv.2210.14537.

Artikel B

Lidmar, J., Vatcher, J., Edelbro, C. & Spross, J. 2023. Estimation of small failure probabilities using the Accelerated Weight Histogram method. *Probabilistic Engineering Mechanics, 74*, 103501.
https://doi.org/10.1016/j.probengmech.2023.103501

Probabilistic Engineering Mechanics 74 (2023) 103501

Contents lists available at ScienceDirect

Probabilistic Engineering Mechanics

journal homepage: www.elsevier.com/locate/probengmech

Estimation of small failure probabilities using the Accelerated Weight Histogram method

Jack Lidmar [a,*], Catrin Edelbro [b], Jessa Vatcher [b], Johan Spross [c]

[a] Department of Physics, KTH Royal Institute of Technology, 106 91, Stockholm, Sweden
[b] Itasca Consultants AB, Aurorum 2, Luleå, 977 75, Sweden
[c] Division of Soil and Rock Mechanics, KTH Royal Institute of Technology, 100 44, Stockholm, Sweden

ARTICLE INFO

MSC:
0000
1111

Keywords:
Structural safety
Rare event simulation
Monte Carlo simulation
Tunnel
Reliability-based design

ABSTRACT

Simulation of rare events, such as small failure probabilities, is a common problem in several scientific and engineering fields. This paper presents a novel Monte-Carlo-based simulation approach for this purpose, called the Accelerated Weight Histogram (AWH) method. The method was originally developed to solve challenging sampling problems in statistical and biological physics, but its algorithm has here been reformulated for estimation of rare event probabilities. The applicability of the method is investigated for a couple of simpler computational examples and for a more advanced practical case, which consists of a rock tunnel stability problem. To estimate the probability of failure of the latter in a realistic manner, a boolean indicator function was used to describe failure, based on a mechanical concept known as unbalanced force ratio. The investigated cases indicate that the AWH method performs well for both simpler limit states and more complex failure definitions.

1. Introduction

Simulation of rare events is a common problem in both physics and engineering, including diverse applications such as free energy calculations, structural safety assessments, and failure probabilities of general technical systems. In this paper, we discuss the simulation of rare events in the context of structural safety assessments; hence, we describe the rare event analysis as an estimation of a probability of failure, which in general can be described by the following integral:

$$P(\mathcal{F}) = \int_{\Omega} \mathbb{1}(x \in \mathcal{F}) \pi(x) dx \tag{1}$$

where $\mathbb{1}(x \in \mathcal{F})$ is an indicator function of failure behavior, and $\pi(x)$ is the joint probability density function of the random variables $x = (x_1, \ldots, x_n) \in \Omega$. The structural behavior is often described by a performance function, $G(x)$, defining the failure event as $\mathcal{F} = \{G(x) \leq 0\}$, though sometimes $G(x)$ is difficult to establish.

Depending on the definition of the failure event and the computational cost to analyze the structural behavior, different methods to assess $P(\mathcal{F})$ can be appropriate. If a tractable analytical and differentiable expression is available for $G(x)$, approximate methods like the first-order or second-order reliability methods provide estimations of $P(\mathcal{F})$ [1–3]. If an analytical expression for $G(x)$ is not directly available or too complex, it can be approximated using a surrogate model, which is used for example in adaptive directional importance sampling [4]. Surrogate models may however have difficulties for complex limit states and in high-dimensional space. For high-dimensional problems, line sampling [5] or asymptotic sampling [6] may be useful alternatives.

Monte Carlo simulation is, however, generally the most robust method, but as the evaluated $P(\mathcal{F})$ generally is small, often in the range 10^{-5} to 10^{-7} for structures, crude Monte Carlo can be unpractical due to its many required evaluations of the model. This method has therefore been improved upon into a number of advanced simulation methods. Importance sampling methods can drastically reduce the number of samples needed, but it is often difficult to construct a suitable importance distribution in more complicated cases. Subset simulation [7,8] solves the problem by introducing a sequence of nested subsets corresponding to ever rarer events, keeping the conditional probabilities large enough to be efficiently sampled using Markov chain Monte Carlo (MCMC). Subset simulation has gained considerable attention over the last decade, see e.g. [9–13], but also some criticism [14].

In this paper, we introduce a new Monte Carlo-based approach to estimate $P(\mathcal{F})$. The method is called the Accelerated Weight Histogram

* Corresponding author.
E-mail address: jlidmar@kth.se (J. Lidmar).

https://doi.org/10.1016/j.probengmech.2023.103501
Received 4 May 2023; Received in revised form 6 July 2023; Accepted 1 August 2023
Available online 9 August 2023

(AWH) method and was originally developed to tackle challenging sampling problems in statistical and biological physics [15]. We have reformulated its algorithm to estimate the probability of rare events, such as failure probabilities of engineering structures. The AWH method uses adaptive importance sampling through a Markov chain that carries out a random walk within a family of distributions that bridge the rare event probability of interest to a known reference. Similar to crude Monte Carlo and subset simulations, the AWH method does not require detailed knowledge about the failure modes of the problem. The potential of the AWH method to estimate failure probabilities was briefly discussed in our recent paper [16], where we found the AWH method to be competitive to subset simulation, especially for hard problems (complex performance functions). In the present paper, we extend this work. In particular, we discuss the ability of the AWH method to handle not only continuous performance functions $G : \Omega \to \mathbb{R}$ crossing some threshold, but also failure described by a boolean indicator function $\mathbb{1}_F : \Omega \to \{0, 1\}$. We note that problems with continuous performance functions can be handled straightforwardly by defining intermediate failure domains $G(\cdot) \le \lambda$, as is done in subset simulation. The boolean case, however, requires a different approach. In Section 5, we present a practical example of a tunnel stability problem, in which failure is described by such a boolean indicator function.

2. The accelerated weight histogram (AWH) method

2.1. General concepts

We begin by describing the AWH method [15] in a general setting, before turning to the particulars of rare event estimation. The AWH method is an adaptive importance sampling Markov chain Monte Carlo (MCMC) method able to sample from a whole family of probability distributions $P(x|\lambda)$ in a single simulation. Here $x \in \Omega$ denotes the (often high-dimensional) stochastic variables that enter the probabilistic model of interest, and $\lambda \in \Lambda$ is a parameter. By transitioning between different λ in a suitable family, the correlation time is often significantly reduced and the sampling of rare regions enhanced. In most cases, the distributions $P(x|\lambda) = Q(x|\lambda)/Z_\lambda$ are known only up to an unknown normalization constant $Z_\lambda = e^{-F_\lambda}$, and may, without loss of generality, be written as

$$P(x|\lambda) = e^{F_\lambda - E_\lambda(x)}, \tag{2}$$

since working with logarithms turns out convenient. This does not prevent sampling from Eq. (2) using Markov chain Monte Carlo, and we assume that such a method is available. As it turns out the normalization constants $Z_\lambda = e^{-F_\lambda}$ will be directly related to the failure probability, and their determination is therefore a central objective of the approach. We may note that the computation of normalization constants have other applications as well. For example, in statistical physics F_λ corresponds to the free energy, while in Bayesian statistics Z_λ corresponds to the evidence, which is useful for model validation and comparisons. The basic idea is now to promote the parameter λ to a dynamical variable and design a Markov chain with a joint equilibrium distribution

$$P(x, \lambda) = \frac{1}{\mathcal{Z}} e^{f_\lambda - E_\lambda(x)}. \tag{3}$$

This can easily be accomplished by alternating MCMC updates of x at fixed λ and updates of λ at fixed x.

The hyperparameters f_λ introduced in Eq. (3) control the marginal distribution of parameters

$$P(\lambda) = \int_\Omega P(x, \lambda) dx = \frac{1}{\mathcal{Z}} e^{f_\lambda - F_\lambda}, \tag{4}$$

where $\mathcal{Z} = \sum_\lambda e^{f_\lambda - F_\lambda}$. The idea is to make this cover an entire range of λ with sufficiently large probability. In order to make the marginal follow a prescribed target distribution π_λ, we need to set $f_\lambda \approx F_\lambda + \ln \pi_\lambda$. Since this depends on the unknown F_λ this is quite nontrivial, but the AWH method accomplishes this by fine tuning f_λ adaptively during the simulation, as described in Section 2.3. When converged, the AWH method will produce both an estimate of the normalization constants $Z_\lambda = e^{-F_\lambda}$ (up to a multiplicative factor), and Monte Carlo samples from Eq. (3), which may be used to compute conditional expectations.

2.2. Parameter moves and hyperparameter updates

In this section, we describe the MCMC moves of λ at fixed x and the adaptive updates of the hyperparameters f_λ. To be practical, the parameter λ is restricted to a discrete set $\Lambda = \{\lambda_k\}_0^M$, and we may use the index k interchangeably with λ_k in the following. Rather than performing only nearest neighbor transitions for the parameter moves $\lambda_{m'} \to \lambda_m$, we use a Gibbs sampler to allow for larger jumps. This allows for a rather fine discretization of λ-values without sacrificing efficiency. Thus, for the current x all $M + 1$ possible moves are considered, and the new state m is drawn with the probability

$$w_m(x) \equiv P(\lambda_m|x) = \frac{e^{f_m - E_m(x)}}{\sum_k e^{f_k - E_k(x)}}. \tag{5}$$

Expectations conditional on λ_k may be estimated from the samples $\{x_i\}_1^n$ as

$$\mathbb{E}\left[A|\lambda_k\right] \approx \bar{A}_k = \frac{\sum_{i=1}^n A(x_i) w_k(x_i)}{\sum_{i=1}^n w_k(x_i)}. \tag{6}$$

The weights $w_k(x)$ are also accumulated in a *weight histogram* W_k. For fixed hyperparameters f_k, the normalized weight histogram will approximate the marginal parameter distribution, $W_k \approx N P(\lambda_k)$, given in Eq. (4), thereby allowing us to estimate the normalization constant $Z_k = e^{-F_k} \approx e^{-f_k} W_k / N\pi_k$, where $N = \sum_k W_k$. In AWH, the hyperparameters are refined iteratively from a sequence of simulations run with different $f_k^{(i)}$ and $\pi_k^{(i)}$. The total accumulated weight histogram $W_k^{(n)} = \sum_{i=1}^n w_k^{(i)}(x_i)$ will then approximately follow $\sum_{i=1}^n \mathcal{Z}_i^{-1} e^{f_k^{(i)} - F_k}$, which may be turned into a convenient recursive update relation

$$W_k^{(n)} = W_k^{(n-1)} + w_k^{(n)}(x_n) \tag{7}$$

$$f_k^{(n+1)} = f_k^{(n)} - \ln\left(\frac{W_k^{(n)}}{W_k^{(n-1)} + \pi_k^{(n)}}\right). \tag{8}$$

Note that the magnitude of the updates to f_k decreases with increasing number of samples, and eventually $f_k^{(n)}$ converges towards $F_k + \ln \pi_k$ (up to a constant) while $W_k \to N\pi_k$.

2.3. The AWH algorithm

The whole algorithm is summarized as follows:

Initialize $m = M$ and $x \sim \pi(x|\lambda_M)$.

Repeat for $n = 1, \ldots, N_{\text{it}}$ (or until the desired accuracy has been reached):

1. Carry out one or more MCMC steps at fixed m:
   (a) Propose a new state x' with probability $q(x'|x)$.
   (b) Accept, i.e. set $x \leftarrow x'$, if $q(x|x')P(x'|\lambda_m)/q(x'|x)P(x|\lambda_m) \ge u$, where $u \sim U[0, 1)$ is a uniform random variate in $[0, 1)$.
   (c) Otherwise set the new state equal to the old one, $x \leftarrow x$.
2. Calculate the weights $w_k(x) \equiv P(k|x)$ for all k, using Eq. (5).
3. Update the weight histogram, $W_k^{(n)} = W_k^{(n-1)} + w_k(x)$.
4. Accumulate weighted averages according to Eq. (6).
5. Choose a new level index m with probability $w_m(x)$.
6. Update the hyperparameters using Eq. (8).

Like many other methods, AWH may only estimate ratios of normalization constants, $Z_k/Z_M = e^{F_M - F_k} \approx e^{f_M - f_k}\pi_k/\pi_M$; absolute normalization constants require a known reference, e.g., $Z_M = 1$.

Before the algorithm starts the hyperparameters f_k must be initialized with a first guess (e.g., $f_k^{(0)} = 0$ if nothing is known), and the weight histogram $W_k^{(0)} = N_{\text{init}}\pi_k$, where N_{init} is a small number, typically of order 1 or M, which quantifies our prior belief in the initial guess. As the simulation goes on the weight histogram will converge towards $N\pi_k$ and the updates will become smaller and smaller $\Delta f_k \sim 1/N \to 0$, while $f_k \to F_k + \ln\pi_k$ when the total number of (effective) samples N grows large. In the early stages of the algorithm the estimates f_k will, however, typically be very poor, which can be diagnosed by a very skewed weight histogram $|W_k - N\pi_k| \gg 0$. When this happens it is recommended to reduce the effective number of samples $N = \sum_k W_k$, e.g. by setting $W_k \leftarrow \min(W_k, cN\pi_k)$ for all k and then $N \leftarrow \sum_k W_k$, where c is a suitable relative tolerance for the acceptable deviations. This prevents large peaks from building up in the weight histogram. A value of $1.25 \le c \le 2$ seems to work well in many cases.

A straightforward generalization of the algorithm is to simulate several systems (sequentially or in parallel), which all contribute to the same weight histogram. Each simulation, below referred to as a "walker", has its own state variables x and level index m, and move around independently from each other, but share W_k and f_k. Besides taking advantage of parallel computing architectures, already a modest number of walkers (~2–4) can in our experience help to speed up the initial convergence towards the target distribution.

2.4. Target distribution

The target distribution $\pi_k \equiv \pi_{\lambda_k}$ is a tuning knob of the algorithm, which in principle may be optimized for convergence. Related to this is the spacing $\delta\lambda_m$ of parameter values, which must be small enough to get a reasonable acceptance probability for the parameter moves. Since the Gibbs sampler [Eq. (5)] can make large jumps along the λ-coordinate a rather dense sequence of λ-values may be used. As long as $\delta\lambda$ is small enough, a nonuniform distribution of levels λ_m will have a similar effect as a nonuniform target distribution π_k, hence we may take $\delta\lambda =$ constant for simplicity. The common naive choice of taking also π_k constant often works well, but is in general suboptimal [17]. Here we opt to choose a target distribution $\pi(\lambda) \propto |dF/d\lambda|$ instead, assuming that F is a monotonic function of λ. A rational for this choice is that it is invariant under nonlinear reparamteterizations $\lambda \mapsto \lambda' = g(\lambda)$, since then $\pi(\lambda')d\lambda' = \pi(\lambda)d\lambda$. This choice is also compatible with the distribution implicitly used in subset simulations, where instead the levels λ_m are chosen adaptively so that $\pi(\mathcal{F}_m)/\pi(\mathcal{F}_{m+1}) = p_0$ is constant with a preset value $p_0 = 0.1 - 0.5$. In practice we have to work with a discrete approximation $\pi_k \propto |\delta F_k|$, e.g. using central differences $[\delta F_k = \frac{1}{2}(F_{k+1} - F_{k-1})$, for $0 < k < M$ and $\delta F_0 = F_1 - F_0$, $\delta F_M = F_M - F_{M-1}]$, and with estimates of the logarithmic normalization constants $F_k \approx f_k - \ln\pi_k$ instead of the exact ones. Since these estimates will be unreliable at the early stages of the simulation it is better to mix in also a (small) uniform component, setting

$$\pi_k = \alpha\frac{1}{M+1} + (1-\alpha)\frac{1}{S}\left|\delta F_k\right|, \tag{9}$$

where $S = \sum_k |\delta F_k|$ is a normalization constant and $0 \le \alpha \le 1$ decreasing as the simulation goes on. In fact it seems helpful to always keep a small uniform component to increase the robustness of the algorithm. For example, we may set $\alpha = \gamma/(\gamma + \min W_k) + \epsilon$, with $\gamma \simeq 10$ to 200 and $\epsilon = 0.01$. For definitiveness, we pick $\gamma = 100$ for the simulations presented below. The adjustment of the target distribution may be carried out between step 1 and 2 in the algorithm, and must be accompanied by the adjustment of the hyperparameters to keep $f_k = F_k + \ln\pi_k$.

3. Rare events

In reliability calculations the goal is to estimate the probability $\pi(\mathcal{F})$ of a rare event $\{x \in \mathcal{F}\}$. In order to adapt the above ideas to a rare event sampling scheme, we only need to specify a suitable family of distributions $P(x|\lambda)$, that interpolates between the target probability

$$\pi(x|\mathcal{F}) = \frac{\mathbb{1}(x \in \mathcal{F})\pi(x)}{\pi(\mathcal{F})} = P(x|\lambda_0) \tag{10}$$

and a known reference $P(x|\lambda_M)$, for instance $\pi(x)$. As discussed, the normalization constant here is just the initially unknown failure probability $\pi(\mathcal{F})$. There is considerable freedom in choosing such a family of distributions and different choices may be preferred in different situations. Quite generally, we take

$$P(x|k) = e^{F_k}\mathbb{1}(x \in \mathcal{F}_k)\pi(x|k), \tag{11}$$

with $\mathcal{F}_0 \equiv \mathcal{F}$ and $\pi(x|0) \equiv \pi(x)$ matching Eq. (10) on one end and $\mathcal{F}_M \equiv \Omega$ and the reference $\pi(x|\lambda_M)$ on the other. The failure probability may then be estimated from the ratio of normalization constants as

$$\pi(\mathcal{F}) = e^{F_M - F_0} \approx e^{f_M - f_0}\frac{\pi_0}{\pi_M}. \tag{12}$$

We begin with a discussion of the common case where the rare event is characterized by some continuous limit state function $G : \Omega \to \mathbb{R}$. Next we discuss the case of a boolean indicator function $\mathbb{1}_{\mathcal{F}} \equiv \mathbb{1}(\cdot \in \mathcal{F}) : \Omega \to \{0, 1\}$.

3.1. Continuous limit state function

Here we assume that failure corresponds to a situation where the limit state function $G(x)$ reaches below a given threshold, arbitrarily set to 0, i.e, $\mathcal{F} = \{G(x) \le 0\}$. In that case it is natural to introduce a nested sequence of failure events $\mathcal{F} \equiv \mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}_M \equiv \Omega$ corresponding to different threshold levels $0 \equiv \lambda_0 < \lambda_1 < \lambda_2 < \cdots < \lambda_M = \infty$, i.e., $\mathcal{F}_k = \{G(x) \le \lambda_k\}$. The event $\mathcal{F}_M$ will then occur certainly, $P(\mathcal{F}_M) = 1$, and will be our reference state. A family of distributions may in this case be defined as

$$P(x|\lambda_k) = e^{F_k}\mathbb{1}(G(x) \le \lambda_k)\pi(x) \tag{13}$$

for $k = 0, \dots, M$.

The joint distribution of x and λ_k (or just k) becomes

$$P(x, \lambda_k) = \frac{1}{\mathcal{Z}}e^{f_k}\mathbb{1}(G(x) \le \lambda_k)\pi(x), \tag{14}$$

while the weights become

$$w_m(x) \equiv P(m|x) = \frac{e^{f_m}\mathbb{1}(G(x) \le \lambda_m)}{\sum_k e^{f_k}\mathbb{1}(G(x) \le \lambda_k)}. \tag{15}$$

3.2. Boolean limit state function

In case the failure region is determined by a boolean function one may proceed in several other ways. One approach is to introduce an approximate performance function that can be used as a proxy for $G(x)$, and proceed along the lines of the previous section. Another approach is to define a family of probability distributions as

$$P(x|\lambda_k) = \frac{\mathbb{1}(x \in \mathcal{F})\pi(x|\lambda_k)}{\pi(\mathcal{F}|\lambda_k)}, \qquad k = 0, 1, \dots, M-1, \tag{16}$$

and $P(x|\lambda_M) = \pi(x|\lambda_{M-1})$. The distributions $\pi(x|\lambda_k)$ must be chosen such that $\pi(x|\lambda_0) \equiv \pi(x)$, and that the rare event becomes much more likely as k increases. An example will be given below in Section 5.3, where the material strength is reduced by a factor of safety $s_k = e^{\lambda_k}$ to make structural failure more probable. The corresponding joint distribution is

$$P(x, \lambda_k) = \frac{1}{\mathcal{Z}}e^{f_k}\mathbb{1}_{\mathcal{F}}(x)\pi(x|\lambda_k), \qquad k = 0, 1, \dots, M-1, \tag{17}$$

$$P(x, \lambda_M) = \frac{1}{\mathcal{Z}}e^{f_M}\pi(x|\lambda_{M-1}), \tag{18}$$

from which the weights $w_m(x) \equiv P(\lambda_m|x) = P(x,\lambda_m)/P(\lambda_m)$ are computed as

$$w_m(x) = e^{f_m}\mathbb{1}_{\mathcal{F}}(x)\pi(x|\lambda_m)/R(x), \qquad m = 0,\dots,M-1, \tag{19}$$

$$w_M(x) = e^{f_M}\pi(x|\lambda_{M-1})/R(x), \tag{20}$$

$$R(x) = e^{f_M}\pi(x|\lambda_{M-1}) + \sum_{k=0}^{M-1} e^{f_k}\mathbb{1}_{\mathcal{F}}(x)\pi(x|\lambda_k). \tag{21}$$

3.3. Monte Carlo moves

Proper sampling of the relevant parameter regions requires efficient Monte Carlo moves also for the random model parameters x at fixed λ. AWH can be paired with any convenient MCMC update that leaves $P(x|\lambda)$ invariant, replacing the Metropolis update [step 1 of the algorithm in Section 2.3]. When $\pi(x|\lambda)$ is a multidimensional standardized normal distribution one may, e.g., use the procedure suggested in [9] or [18], drawing each proposal x_i' from $q(\cdot|x) = N(\sqrt{1-\delta^2}x_i, \delta^2)$, where $0 < \delta \le 1$ is a suitably chosen step length, and accepting if $x' \in \mathcal{F}_m$. This will be used frequently in the examples that follow. Another possibility is to use Hybrid/Hamiltonian Monte Carlo updates.

Often the level $m = M$ is such that it is possible to sample directly from $P(x|\lambda_M)$ and then that is obviously preferable.

4. Computational examples

The AWH method has been tested out on a number of different benchmark problems. In a previous publication [16] we applied it to two model problems, a normal distribution with linear limit state function and a less trivial fiber-bundle model. These examples are briefly discussed below together with a couple of other benchmark problems of varying complexity. A more realistic problem is studied in Section 5.

We have also compared with subset simulation, which is a well-established and highly performant method for rare event simulation [7,8]. Subset simulation is a sequential Monte Carlo method that evolves a population of samples towards the rare failure region via a combination of selection, cloning, and MCMC. We used a fraction $p_0 = 0.2$ (or sometimes 0.1) of the sample population as seeds for subsequent levels. The population size was adjusted so that the number of function evaluations of $G(\cdot)$ were the same for both subset and AWH simulations, to make comparisons fair and easy.

A robust way to assess the accuracy of the methods is to perform many independent runs of the simulations and apply standard statistical methods to the resulting estimates. We compute the coefficient of variation (CV) and the root-mean-square (RMS) deviation from exact results, where available.

The AWH estimate is asymptotically unbiased, implying that any bias is negligible compared to the statistical error for long simulations. The bias comes from the initial transient as the histogram builds up and the Markov chain equilibrates, and decays inversely with simulation time.

4.1. Sum of log-normal random variables

As a first illustration of how the AWH method works, consider a model where failure occurs if the sum of n log-normal random numbers exceeds a given threshold. The limit state function is given by [19]

$$G(x) = n + b\tilde{\sigma}\sqrt{n} - \sum_{i=1}^{n} y_i, \tag{22}$$

where y_i are log-normal random numbers with mean 1 and standard deviation $\tilde{\sigma} = 0.2$. The MCMC updates are performed in standard normal space, i.e., with $x_i \sim N(0,1)$ and $y_i = \exp(\mu + \sigma x_i) \sim LN(\mu,\sigma^2)$. There is no exact result to compare with, but for large n the sum of log-normal random variables will approach a normal one, hence the failure probability will approximately be $\Phi(-b)$ in that limit, where Φ is the standard normal cumulative distribution. The convergence to a normal distribution is rather slow, however, due to the heavy tails of the log-normal one. We choose the threshold $b = 5$, which gives $\Phi(-5) = 2.8665 \times 10^{-7}$, and a moderate $n = 50$ for which deviations from normal may be seen. The actual failure probability in this case was estimated by the AWH simulations to be 1.36×10^{-6} after 100 000 function evaluations of $G(\cdot)$. The coefficient of variation of the final estimate of $P(\mathcal{F})$ was estimated to 0.1 ± 0.01 from 200 repetitions of the simulation. Subset simulations for the same setup also gave 0.1 ± 0.01.

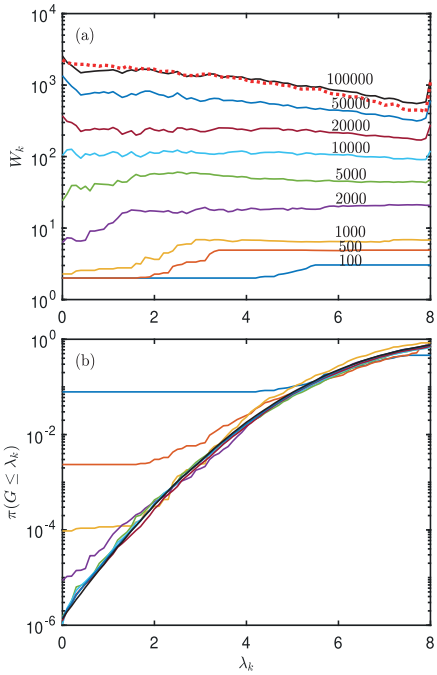


Fig. 1. (a) Weight histogram and (b) estimate of failure probability $\pi(G \le \lambda_k)$ after 100, 500, …, 100 000 iterations of an AWH simulation of the log-normal distribution example. The dotted line in (a) shows the target distribution $N\pi_k$.

The AWH simulations used 82 levels $\lambda_k \in \{0, 0.1, 0.2, \dots, 8.0, \infty\}$ and 4 walkers. The MCMC updates of x_i at fixed λ_m used proposals $q(\cdot|x_i) = N(\sqrt{1-\delta^2}x_i, \delta^2)$, with $\delta = 0.6$, which were accepted if $G(x) \le \lambda_m$.

Fig. 1(a) shows how the weight histogram grows during the simulation, while (b) shows the corresponding evolution of the failure probability estimates as function of λ. For small number of iterations the histogram is very skewed and does not cover the whole range of λ, leading to very poor estimates of $\pi(\mathcal{F}_k)$. As soon as the whole range of λ is covered the failure probability estimates start to become reasonable. As the simulation goes on the weight histogram eventually converges towards the target distribution $N\pi_k$ chosen as in Eq. (9).

4.2. Sum of normal random variables

A useful test case, similar to the previous but amenable to analytical calculations, can be formulated using a sum of normal random variables. We consider a limit state function $G(x) = b - n^{-1/2}\sum_{i=1}^{n} x_i$, where $x_i \sim N(0,1)$. Since the sum of normally distributed variables is again normal the exact failure probability is $P(\mathcal{F}) = \Phi(-b)$. We set $n = 2$ and $b = 6$, which gives $P(\mathcal{F}) \approx 0.9866 \times 10^{-9}$. A single AWH simulation using 100 000 iterations, 4 walkers, and $\lambda \in \{0, 0.1, 0.2, \dots, 8.0, \infty\}$ resulted in an estimate $P(\mathcal{F}) \approx 0.85 \times 10^{-9}$. To compute errors, we repeated the

simulation 200 times, which gave a relative root-mean-square (RMS) deviation from the exact result of 0.21 ± 0.01 (using bootstrap to get the standard error). The coefficient of variation was also estimated to 0.21 ± 0.01.

For comparison we also ran 200 independent subset simulations, which resulted in a relative RMS and coefficient of variation both equal to 0.17 ± 0.01. We can see that both AWH and subset simulation give good approximations to the exact result in this test case, with similar errors.

4.3. Parallel system

Another example that is analytically tractable is a system of n redundant components. Each component can fail independently with a probability p, and system failure corresponds to all of them failing, which occurs with a probability $P(\mathcal{F}) = p^n$. For the concrete realization of the model, we assume that a component i fails if its capacity C is exceeded by a random load r_i, where the loads have log-normal distributions $LN(0,1)$. Thus we set $r_i = \exp(x_i)$, $x_i \sim N(0,1)$ and define the failure domain as $\mathcal{F} = \{r_i \geq C\} = \{x_i \geq \ln C\}$. A continuous limit state function may be introduced as

$$G(x) = C - \min_i \exp(x_i). \tag{23}$$

The exact failure probability is $P(\mathcal{F}) = (1 - \Phi(\ln C))^n$. We set $C = 1$ and $n = 20$, which gives $P(\mathcal{F}) = 2^{-20} \approx 0.954 \times 10^{-6}$.

We study the model using AWH and subset simulations, employing the same type of MCMC attempts as in Section 4.1. The estimated failure probability from a single AWH simulation with 100 000 iterations, 4 walkers, and $\lambda \in \{0, 0.05, \dots, 0.9, \infty\}$ was $P(\mathcal{F}) \approx 0.978 \times 10^{-6}$. Repeating 200 independent simulations gave a relative RMS deviation from the exact result of 0.26 ± 0.014. Subset simulations for the same setup, on the other hand, gave a relative RMS 0.34 ± 0.018.

4.4. Network flow model

This example consists of a network H of nodes connected via links. Each link has a random capacity to carry a maximal amount of flow. The capacities are assumed to follow independent log-normal distributions $y_i = \exp(0.5x_i) \sim LN(0, 0.5^2)$. Failure occurs if the maximal network flow between a source and a sink exceeds a given threshold b, giving a limit state function $G(x) = b - \text{maxflow } H$. For concreteness we study a network of nodes on a 20×20 square lattice, see the inset of Fig. 2, with source and sink nodes indicated by green and red, respectively. The threshold is set to $b = 10$. AWH simulations are performed using $M + 1 = 26$ levels $\lambda_k \in \{0, 0.25, 0.5, \dots, 6, \infty\}$, and 4 walkers. The MCMC moves were the same as in the previous example. The failure probability as function of λ is shown in Fig. 2, with the weight histogram after 100 000 samples shown in the inset. The estimated failure probability was 3.5×10^{-9} in this case. A coefficient of variation of 0.30 ± 0.02 was estimated from 200 independent repetitions of the simulation. As a comparison, subset simulations gave a value of 0.43 ± 0.03.

4.5. The fiber bundle model

As a final, more challenging example, we discuss failure in the Fiber Bundle Model (FBM), studied in Ref. [16]. This model was originally introduced to describe the strength of textiles, but has since found use in a wide range of applications, including fracture, wire cables, earthquakes, and landslides [20–22]. It consists of N elastic strings or fibers connected in parallel to a common load. The fibers have identical spring constants κ, but break at different random strains x_i, which are independent and identically distributed according to $\pi(x_i)$. When the load slowly increases from zero to a final value L, the weaker fibers will break and the stress will be redistributed among the remaining ones. This can cause additional failures and so on. In fact, for large N

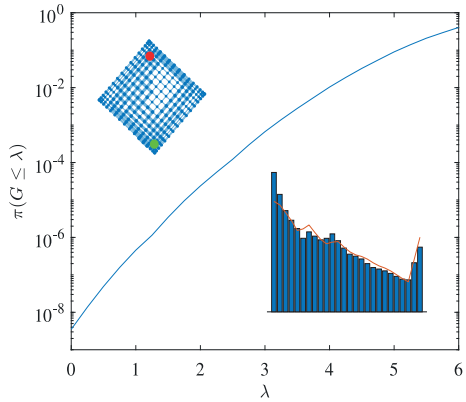


Fig. 2. Failure probability $\pi(G \leq \lambda)$ for the network flow model as function of threshold λ. The upper left inset shows the network with the source and drain indicated in green and red. The lower right inset shows the corresponding weight histogram (blue bars) and target distribution (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the failure of the structure occurs via a series of avalanches or bursts, which follow a powerlaw distribution on approaching the critical load L_c at which all fibers have snapped. The total force as function of the extension ϵ is

$$F(\epsilon) = \sum_{i=1}^{N} \kappa\epsilon \mathbb{1}(\epsilon \leq x_i). \tag{24}$$

The effect of the indicator function is to include only the intact strings with extension less than their threshold, $\epsilon \leq x_i$. On average the force becomes $\overline{F}(\epsilon) = \mathbb{E}[F(\epsilon)] = N\kappa\epsilon(1 - \Pi(\epsilon))$, where $\Pi(\cdot)$ is the cumulative distribution of x_i. The variance is $\text{Var } F(\epsilon) = N\kappa^2\epsilon^2\Pi(\epsilon)(1 - \Pi(\epsilon))$. In the limit $N \to \infty$ the relative fluctuations around the mean tend to zero so that the structure almost surely holds for $L < L_c$ and fails for $L > L_c$, where $L_c = \max_\epsilon \overline{F}(\epsilon)$. For concreteness we set $\kappa = 1$ and assume that the thresholds are uniformly distributed, $x_i \sim U[0,1)$. Then $\Pi(x) = x$ and one finds a maximum force $L_c = N/4$ corresponding to an extension $\epsilon_c = 1/2$.

Failure below the threshold can, for finite N, occur as the result of a rare fluctuation, and may be described by a limit state function [16]

$$\begin{aligned} G(\mathbf{x}) &= \max_\epsilon F(\epsilon) - L \\ &= \kappa \max_j x_j \sum_{i=1}^{N} \mathbb{1}(x_j \leq x_i) - L. \end{aligned} \tag{25}$$

For the MCMC updates at fixed λ we use simple Metropolis attempts in which the threshold x_i of a randomly selected fiber i is replaced by a uniform random variate in $[0,1)$.

Let us first examine the case $L = 200 < 250 = L_c$ studied in Ref. [16]. An AWH simulation with levels $\lambda \in \{0, 1, \dots, 60, \infty\}$ was executed for 200 000 iterations, after which the weight histogram appeared reasonably well converged, and gave an estimate of $P(\mathcal{F}) \approx 1.7 \times 10^{-13}$. Continuing up to 20 million iterations refined this to 1.4×10^{-13}.

In comparison, subset simulations generally encountered difficulties for the FBM [16]. Initially, we performed subset simulations with population size varying between 1000 and 100 000 samples, using a fraction $p_0 = 0.1$ as seeds for subsequent levels. However, they did not converge unless the number of MCMC updates at each level was increased. With a 10-fold increase the subset simulations were able to reproduce the AWH simulation, but only with a much larger number $\gtrsim 1\,000\,000$ of model evaluations. The low efficiency of the MCMC updates may partly be blamed for this.

To compare errors we also study a less demanding case with a larger load $L = 220$ and therefore a higher probability of failure. The levels were chosen as $\lambda \in \{0, 1, 2, \ldots, 40, \infty\}$. A nearly error-free estimate $P(F) \approx 4.8 \times 10^{-6}$ was obtained from a very long AWH simulation after 50×10^6 iterations, which we use as a reference value in the absence of an exact calculation. A single shorter AWH simulation using 500 000 iterations gave an estimate $P(F) \approx 4.2 \times 10^{-6}$, with the coefficient of variation and the relative RMS from the reference value both equal to 0.21 ± 0.01, obtained from 200 independent runs.

For subset simulations using a sample population of 10 000, $p_0 = 0.1$, and 100 (instead of the standard $1/p_0 = 10$) MCMC updates per seed and level, the RMS relative error and coefficient of variation were 1.1 ± 0.1 from 200 repetitions. The computational effort was the same in both cases, requiring 500 000 model evaluations of $G(\cdot)$ per simulation. Thus, for this test case AWH significantly outperformed subset simulation, with a coefficient of variation five times smaller.

5. Application example: Tunnel

5.1. Case description

We have also applied the AWH method to an illustrative practical example: the stability of a tunnel excavated in rock. The tunnel is 10 m wide and located at a depth $h = 300$ m below surface. The rock mass is very weak and of such character that it is not possible to test with standard procedures in lab, so the rock mass properties are very uncertain. The applied random variables of the rock mass (x_i) are summarized in Table 1 and introduced in the following. The conditions are comparable in magnitudes to what Hoek and Brown [23] describes as "poor quality rock mass under high stress". All random variables are for convenience in this illustrative example assumed independent and log-normal. The behavior of the AWH method is investigated in three different scenarios, which have been assigned slightly different properties and coefficients of variation. Following the same underlying assumptions as in Spross et al.'s [24] detailed probabilistic design example of a tunnel, the probability distributions are meant to reflect the epistemic uncertainty in the parameters' spatial average. Considering the rock mass to be homogeneous in this manner is reasonable when the considered rock mass volume is larger than the Representative Elementary Volume of a fractured rock mass [25], which we for simplicity assume acceptable in this calculation example.

For the rock mass behavior, the Mohr–Coulomb shear failure criterion is used, as is common in tunnel design. Here it is applied with peak strength for its input parameter values cohesion c, friction coefficient $\tan\phi$, both of which are modeled as random variables. The rock mass density (2650 kg/m^3), dilation angle (2°), and tensile strength (200 Pa) are assumed deterministic. The in-situ stresses in the rock mass are caused by tectonic stresses and gravitational stresses from the overburden. The three principal stresses are assumed directed horizontally (σ_x, σ_y) and vertically (σ_z) in relation to the tunnel axis. The largest horizontal stress σ_x (in terms of mean value) is assumed to be oriented perpendicular to the axis of the tunnel. The magnitudes of the principal stresses are modeled with considerable uncertainty, as is indicated by major field measurement campaigns, e.g., [26,27] (even for the vertical stress, which in practice often is estimated to be $\sigma_z = \rho g h$ in deterministic analyses). Both horizontal and vertical stresses are assumed to increase with depth, which is described by a gradient (σ'_x, σ'_y, σ'_z). The average values of the stress state are assumed to be in accordance to that of the central parts of Stockholm [26]. The orientation of the principal stresses is however assumed deterministic.

The rock around the tunnel is reinforced by both bolting and surface reinforcement, the properties of which are assumed deterministic. The bolt reinforcement (25 mm) has a spacing of 2 m and is simulated in the model with an ideal plastic material model, having a yield tensile capacity of 214 kPa and tensile failure strain of 4.3%. The surface reinforcement consists of 0.1 m shotcrete, simulated with an elastic material model, having a Young's modulus of 16 GPa and Poisson's ratio of 0.25.

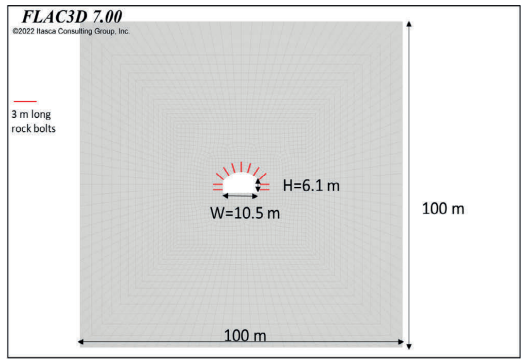


Fig. 3. Cross sectional view of used FLAC3D model of the rock mass around the 10 m-wide tunnel.

5.2. Numerical stress analysis models

The numerical stress analysis models were created in Itasca's FLAC3D finite difference continuum software [28]. Although run in a three-dimensional environment, a quasi-3D model was used, utilizing plane-strain symmetry conditions to reduce computational time. Once a set of variables were selected using the AWH method, Python was used to build and run the specific model in FLAC3D. Output from FLAC3D was then reported back to the AWH algorithm.

The models used in the simulations consisted of a standard tunnel cross-section geometry (Fig. 3). Zone size near the tunnel was set to 1 m, gradually increasing towards the outer boundaries. The model thickness (out-of-plane) was 1 m, chosen to make near-square-shaped zones optimal for stress and plasticity calculations. Roller boundaries were used on all sides of the model. Models were run in small strain (i.e., the grid point positions are not physically updated) to allow for comparison of the presented methodology to Itasca's built-in factor of safety methodology during the simulations.

5.3. Applied failure definition

Failure of a rock engineering structure is not easily modeled. Several different failure definitions have been proposed and used over the last decades, e.g., [29–33]. A key issue is that failure of the rock mass itself, i.e., attained failure envelope, does not necessarily equal failure of the structure, for example represented by block fall or tunnel collapse. To make the calculated failure probabilities comparable to target failure probabilities in design codes, a failure definition simulating structural instability has to be applied. In this paper, failure through instability was evaluated by studying whether the model reaches equilibrium or not. In FLAC3D, instability is assessed through the model's change in kinetic energy, which is measured by an "unbalanced force ratio" [34]. If the unbalanced force has fallen below a required level, the model is considered to be in equilibrium (stable); if it has not fallen below this level or even increases, the model represents instability failure.

For most parameters in the stable domain, the FLAC3D calculations will converge to an equilibrium within 10 000 iterations. However, for parameters close to the failure domain the convergence can be much slower. Therefore, many iterations of the unbalanced force ratio can be necessary in order not to misclassify stable parameter points as failed ones. Initially, we considered an unbalanced force ratio above 10^{-5} after 10 000 steps to meet the failure criterion. This, however, resulted in a rather blurred boundary of the failure domain. Since the AWH algorithm will concentrate its effort on those borderline cases and extending the number of iterations by a large amount is costly, an early exit strategy is employed using the following steps.

Table 1
Material parameters. All parameters are assumed to follow log-normal distributions with mean and coefficient of variations CV given in the table.

			Mean	Case 1 CV	Case 2 CV	Case 3 CV
Rock mass properties						
Young's modulus	E	[Pa]	2×10^9	0.5	0.25	0.25
Poisson ratio	ν		0.25	0.1	0.1	0.1
Cohesion	c	[Pa]	0.5×10^6	0.4	0.4	0.4
Friction coefficient	$\tan\phi$		$\tan 33°$	0.4	0.4	0.4
In-situ stress conditions						
Horizontal stress, perpendicular	σ_x	[Pa]	25.1×10^6	0.2	0.3	1.5
Gradient of σ_x	σ'_x	[Pa/m]	0.082×10^6	0.15	0.15	0.25
Horizontal stress, parallel	σ_y	[Pa]	11.3×10^6	0.2	0.3	1.5
Gradient of σ_y	σ'_y	[Pa/m]	0.033×10^6	0.15	0.15	0.25
Vertical stress	σ_z	[Pa]	6.5×10^6	0.12	0.12	0.5
Gradient of σ_z	σ'_z	[Pa/m]	0.026×10^6	0.12	0.12	0.25

First 5000 FLAC3D iterations are performed and if the remaining unbalanced force is below a threshold 10^{-5} the structure is classified as stable. If not, the calculation is extended up to a maximum of 20 000 iterations, while the unbalanced force ratio is checked against the same threshold every 1000 iterations. If the remaining unbalanced force ratio is still above the threshold, but decreasing by more than 2% since the previous check, the calculation is allowed to continue up to a maximum of 100 000 iterations while checking the unbalanced force ratio every 1000 steps. As soon as a check gives an unbalanced force ratio less than the threshold the FLAC3D calculation is terminated and the structure is classified as stable. This procedure avoids many false positives and resulted in a much clearer boundary between the stable and failed regions.

5.4. AWH simulations

For given set of input parameters, the FLAC3D model simulation will determine whether failure occurred or not, according to the failure definition introduced above. Since failure in this case is described by a boolean function $\mathbb{1}_{\mathcal{F}}(\cdot)$, we use the strategy outlined in Section 3.2, using a family of distributions of the form in Eq. (16). At this point, an appropriate parameter λ in the base probability distribution $\pi(x|\lambda)$ is needed, such that failure becomes increasingly likely for large λ. This choice may be done in many different ways, but we choose to implement this using Dawson et al.'s [34] concept of trial safety factors, which they define in their calculation of the unbalanced force ratio. In practice, this implies a reduction of the cohesion c and friction coefficient $\tan\phi$ by a factor

$$s = \exp(\lambda) \geq 1. \tag{26}$$

For the implementation it is convenient to be able to map the variables x_i back and forth to standard normal ones z_i. Thus, we define

$$z_i \sim N(0,1) \tag{27}$$

$$y_i = \exp(\mu_i + \sigma_i z_i) \sim LN(\mu_i, \sigma_i^2) \tag{28}$$

$$x_i = \begin{cases} y_i/s, & i \in \{c, \tan\phi\} \\ y_i, & \text{otherwise,} \end{cases} \tag{29}$$

where μ and σ are the location and scale parameters of the lognormal distribution, which are obtained from the mean m and coefficient of variation CV listed in Table 1 as $\mu = \ln m - \sigma^2/2$ and $\sigma^2 = \ln(1 + CV^2)$. The effect of re-scaling by a factor s thus amounts to shifting the location parameter μ by $\lambda = \ln s$, resulting in a distribution for the rescaled parameters (c and $\tan\phi$) equal to $LN(\mu_i - \lambda, \sigma_i^2)$. Denoting the base distribution for parameter x_i by $\pi_i(x_i|\lambda)$ the joint distribution becomes

$$P(x, \lambda_m) = \frac{1}{\mathcal{Z}} e^{f_{\lambda_m}} \mathbb{1}(x \in \mathcal{F}_m) \prod_i \pi(x_i|\lambda_m), \tag{30}$$

where $\mathcal{F}_m = \mathcal{F}$ for $m < M$ and $\mathcal{F}_M = \Omega$.

The Monte Carlo updates of x at fixed λ_m may be carried out by mapping x to standard normal space z and propose a new state $z'_i \sim N(\sqrt{1-\delta^2} z_i, \delta^2)$ for all i. The new state is transformed back to lognormal space and rescaled, and then accepted, $x \leftarrow x'$, if the failure criterion is met for the new x'. Otherwise, it is rejected, i.e., x remains in the previous state. The step length δ may be adapted during simulation for each level m to give a reasonable acceptance rate, e.g., 0.234. The restriction $0.2 \leq \delta \leq 0.9$ is also imposed. The level at $m = M$ is special in that the failure criterion need not be evaluated, and the new state can be drawn directly from $\pi(x|\lambda_M)$ using Eqs. (27)–(29).

The Monte Carlo updates of the level λ_m at fixed x in AWH use a Gibbs sampler, where the new level is selected from all $M+1$ possible values with a probability $w_m(x)$ given in Eqs. (19). For the specific case discussed here, where only the distributions of c and $\tan\phi$ depend on λ, this simplifies to

$$w_m(x) = \frac{1}{R(x)} e^{f_m - \frac{1}{2\sigma_c^2}(\ln c + \lambda_m - \mu_c)^2 - \frac{1}{2\sigma_\phi^2}(\ln\tan\phi + \lambda_m - \mu_\phi)^2} \tag{31}$$

if $x \in \mathcal{F}$, otherwise $w_m(x) = \delta_{m,M}$. Here $R(x)$ is a normalization constant such that $\sum_k w_k(x) = 1$.

The selection of levels $s_m = \exp(\lambda_m)$ is not very critical, and after a little experimentation we settled for a sequence $s_m \in \{1, 1.1, 1.2, 1.3, 1.4, 1.6, 1.8, 2, 2.1, 2.25, 2.5, 2.75, 3, 3.5, 4, 4.5, 5, 6, 7, 8, 8\}$. The target distribution π_m is chosen as in Eq. (9). When converged, the AWH simulations will provide estimates of the failure probability $\pi(\mathcal{F}|s_m)$ also at the intermediate scale factors $s_m > 1$. These may be interpreted as the probability $\pi(s(y) < s_m)$ that the actual factor of safety $s(y) = \inf\{s > 0 : x(y,s) \in \mathcal{F}\}$ is smaller than s_m, where $x(y,s)$ refers to the rescaling in Eq. (29). This interpretation is based on the premise that $y \in \mathcal{F}$ implies that $x(y,s) \in \mathcal{F}$ for all $s \geq 1$, i.e., that reducing the material strength cannot increase the stability of the structure.

5.5. Results

The AWH simulations were run as described in Section 5.4, using 4 simultaneous walkers. Fig. 4 shows the estimated failure probability as a function of the rescaling factor s for the three different cases in Table 1. The number of FLAC3D model runs were 10 000, in each case. The actual failure occurs at $s = 1$ in these plots, giving the failure probability estimates $\pi(\mathcal{F}) = 7.2 \times 10^{-7}, 1.6 \times 10^{-6}$, and 8.3×10^{-6}, respectively. The insets in Fig. 4 show the accumulated weight histograms during the AWH simulations, which are not far from the target distribution (red curves). These are useful to monitor that all s_m get sufficiently sampled.

In Fig. 5 all samples produced by the simulation are shown, projected to the $c - \tan\phi$ plane, as the cohesion c and friction angle ϕ typically are the most important parameters for the stability of tunnel structures. The red points correspond to failure and the green to the stable situation being reported from the model run. In case 1 and

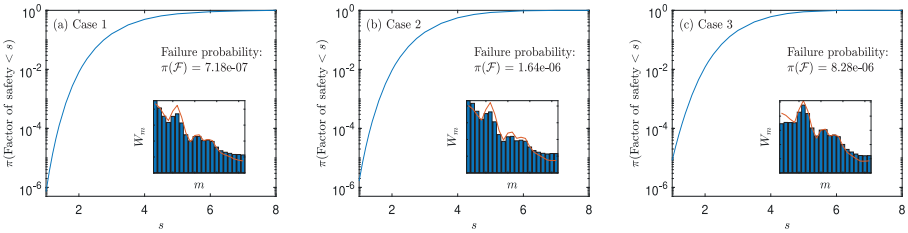


Fig. 4. Probability of failure of the tunnel as function of the scaling factor s. The actual failure corresponds to $s = 1$. The whole curve may be interpreted as the probability that the factor of safety is less than s. The insets show the weight histograms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

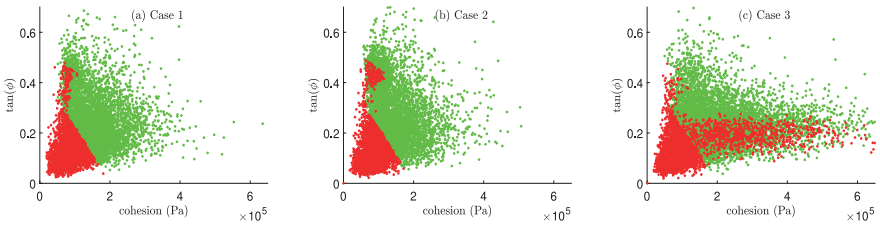


Fig. 5. Model parameters generated by the AWH simulations, projected to the c–$\tan\phi$ plane, for the three test cases. Red points correspond to failure and green to stable parameter values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2, the failure boundary is very clear in this projection, with some deviations visible at low cohesion. Case 3, however, is run with much more uncertainty in the stress state, and the effect of this is seen as failures occurring also for larger values of c.

To learn about the relative importance of the material parameters we may study the sensitivity with respect to variations in their mean and coefficient of variation. One such sensitivity measure is given by

$$\frac{\partial \ln \pi(\mathcal{F})}{\partial \ln \alpha} = \mathbb{E}\left[\frac{\partial \ln \pi(x)}{\partial \ln \alpha}\middle|\mathcal{F}\right], \tag{32}$$

where α is the mean or CV. The conditional expectations are evaluated using Eq. (6), and plotted in Fig. 6. The cohesion and friction coefficient are the most influential variables, particularly in cases 1 and 2, but in case 3 the horizontal stress and the Poisson ratio are also becoming important. This is fully consistent with the shape of the failure boundary seen in Fig. 5.

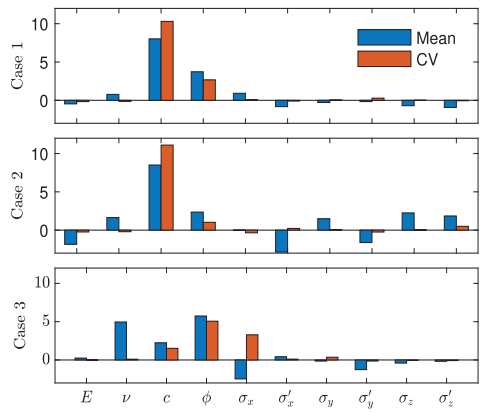


Fig. 6. Sensitivity of the failure probability to perturbations in the mean and coefficient of variation. Blue and red bars correspond to $-\partial \ln \pi(\mathcal{F})/\partial \ln$ Mean and $\partial \ln \pi(\mathcal{F})/\partial \ln$ CV, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Concluding remarks

This paper introduced the Accelerated Weight Histogram (AWH) method for estimation of rare event probabilities, such as small failure probabilities of structures and other technical systems. The performed computational examples in this paper and our briefer earlier account [16] illustrate the method's flexibility in handling efficiently both non-trivial limit states and very small failure probabilities ($\lesssim 10^{-6}$) in a variety of problems, including both continuous and boolean limit state descriptions. The boolean formulation, i.e. $\mathbb{1}_{\mathcal{F}} : \Omega \to \{0, 1\}$, is useful when the explicit formulation of a continuous limit state function is unavailable. This situation is common for example in large-scale instability problems in geotechnical engineering, like tunnel collapse or failure of engineered and natural slopes in soil and rock. In this context, it is worth pointing out that subset simulations need a continuous limit state function, though more general sequential Monte Carlo methods may be used as an alternative.

We have compared AWH with subset simulations for different model problems of varying complexity. These include sums of normal or log-normal random variables, parallel systems with redundancy, and the Fiber Bundle Model with more complicated interactions. We found the performance of AWH to be comparable or sometimes surpassing subset simulations. The AWH method performed particularly well on the Fiber Bundle Model, where many transitions over the parameter range was necessary to reach convergence, and where the subset simulations either failed or required many more model evaluations. A possible explanation for the issues with subset simulation in this test case could be the low efficiency of the MCMC updates used there, which leads to a slow decay of time correlations. Efficient MCMC updates of the model parameters x at fixed λ are helpful also for AWH, but apparently not as critical as for subset simulations.

Successful application of the AWH method to a particular problem relies on finding a good family of intermediate distributions $P(x|\lambda_k)$ bridging the rare failure event to a known reference. This can be done using a suitable performance function $G : \Omega \to \mathbb{R}$, when applicable, or by deforming the probability distributions as in the tunnel example in Section 5. For complex models this can be nontrivial, since some insight

into the failure mechanisms of the problem is often required. Analogous considerations apply to the choice of limit state function G in subset simulations, where a poor choice can severely hamper the efficiency and even prevent convergence [14]. AWH may be similarly affected by a poor choice of G or the family $P(x|\lambda_k)$, although it arguably has more flexibility.

One merit of the AWH method is that problematic cases can be easily diagnosed by monitoring the weight histogram collected during the simulation. If the weight histogram is very skewed or far from the chosen target distribution, this is a clear sign that the sampling is insufficient. If this happens one can either continue the simulation further to collect more samples or choose a better suited family of interpolating distributions.

Future research may be directed to further optimize the target distribution to speed up the convergence and to combine the AWH method with surrogate models in order to reduce the number of computationally demanding model runs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article

Acknowledgments

The presented research was funded and supported by the Rock Engineering Research Foundation (BeFo), Sweden [grant no. 424]. The research was conducted without involvement of the funding source. Joel Andersson, Itasca Consultants AB, is acknowledged for assisting with the numerical model in the tunnel example.

References

[1] A.M. Hasofer, N.C. Lind, Exact and invariant second-moment code format, J. Eng. Mech. Div. 100 (1) (1974) 111–121, http://dx.doi.org/10.1061/jmcea3.0001848.
[2] R. Rackwitz, B. Flessler, Structural reliability under combined random load sequences, Comput. Struct. 9 (5) (1978) 489–494, http://dx.doi.org/10.1016/0045-7949(78)90046-9.
[3] A.D. Kiureghian, H.-Z. Lin, S.-J. Hwang, Second-order reliability approximations, J. Eng. Mech. 113 (8) (1987) 1208–1225, http://dx.doi.org/10.1061/(ASCE)0733-9399(1987)113:8(1208).
[4] F. Grooteman, An adaptive directional importance sampling method for structural reliability, Probab. Eng. Mech. 26 (2) (2011) 134–141, http://dx.doi.org/10.1016/j.probengmech.2010.11.002.
[5] H.J. Pradlwarter, G.I. Schueller, P.S. Koutsourelakis, D.C. Charmpis, Application of line sampling simulation method to reliability benchmark problems, Struct. Saf. 29 (3) (2007) 208–221, http://dx.doi.org/10.1016/j.strusafe.2006.07.009.
[6] C. Bucher, Asymptotic sampling for high-dimensional reliability analysis, Probab. Eng. Mech. 24 (4) (2009) 504–510, http://dx.doi.org/10.1016/j.probengmech.2009.03.002.
[7] S.-K. Au, J.L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probab. Eng. Mech. 16 (4) (2001) 263–277, http://dx.doi.org/10.1016/S0266-8920(01)00019-4.
[8] S.-K. Au, Y. Wang, Engineering Risk Assessment with Subset Simulation, John Wiley & Sons, Singapore, 2014, pp. 1–315, http://dx.doi.org/10.1002/9781118398050.
[9] I. Papaioannou, W. Betz, K. Zwirglmaier, D. Straub, MCMC algorithms for subset simulation, Probab. Eng. Mech. 41 (2) (2015) 89–103, http://dx.doi.org/10.1016/j.probengmech.2015.06.006.
[10] S.-K. Au, On MCMC algorithm for subset simulation, Probab. Eng. Mech. 43 (2016) 117–120, http://dx.doi.org/10.1016/j.probengmech.2015.12.003.
[11] W. Betz, I. Papaioannou, J.L. Beck, D. Straub, Bayesian inference with subset simulation: Strategies and improvements, Comput. Methods Appl. Mech. Engrg. 331 (2018) 72–93, http://dx.doi.org/10.1016/j.cma.2017.11.021.
[12] M. Rashki, SESC: A new subset simulation method for rare-events estimation, Mech. Syst. Signal Process. 150 (2021) 107139, http://dx.doi.org/10.1016/j.ymssp.2020.107139.
[13] J. Chan, I. Papaioannou, D. Straub, An adaptive subset simulation algorithm for system reliability analysis with discontinuous limit states, Reliab. Eng. Syst. Saf. 225 (2022) 108607, http://dx.doi.org/10.1016/J.RESS.2022.108607.
[14] K. Breitung, The geometry of limit state function graphs and subset simulation: Counterexamples, Reliab. Eng. Syst. Saf. 182 (2019) 98–106, http://dx.doi.org/10.1016/j.ress.2018.10.008.
[15] J. Lidmar, Improving the efficiency of extended ensemble simulations: The accelerated weight histogram method, Phys. Rev. E 85 (5) (2012) 056708, http://dx.doi.org/10.1103/PhysRevE.85.056708.
[16] J. Lidmar, J. Spross, J. Leander, Accelerated weight histogram method for rare event simulations, in: Proceedings of the 13th International Conference on Structural Safety and Reliability (ICOSSAR 2021-2022), China, Shanghai, 2022, http://dx.doi.org/10.48550/arXiv.2210.14537.
[17] V. Lindahl, J. Lidmar, B. Hess, Riemann metric approach to optimal sampling of multidimensional free-energy landscapes, Phys. Rev. E 98 (2) (2018) 023312, http://dx.doi.org/10.1103/PhysRevE.98.023312.
[18] S.-K. Au, E. Patelli, Rare event simulation in finite-infinite dimensional space, Reliab. Eng. Syst. Saf. 148 (2016) 67–77, http://dx.doi.org/10.1016/j.ress.2015.11.012.
[19] P. Koutsourelakis, H. Pradlwarter, G. Schuëller, Reliability of structures in high dimensions, part I: algorithms and applications, Probab. Eng. Mech. 19 (4) (2004) 409–417, http://dx.doi.org/10.1016/j.probengmech.2004.05.001.
[20] F.T. Peirce, The weakest link theorems on the strength of long and of composite specimens, J. Text. Inst. Trans. 17 (7) (1926) T355–T368, http://dx.doi.org/10.1080/19447027.1926.10599953.
[21] H. Daniels, The statistical theory of the strength of bundles of threads. I, Proc. R. Soc. Lond. Ser. A Math. Phys. Sci. 183 (995) (1945) 405–435, http://dx.doi.org/10.1098/rspa.1945.0011.
[22] S. Pradhan, A. Hansen, B.K. Chakrabarti, Failure processes in elastic fiber bundles, Rev. Modern Phys. 82 (1) (2010) 499–555, http://dx.doi.org/10.1103/RevModPhys.82.499.
[23] E. Hoek, E.T. Brown, Practical estimates of rock mass strength, Int. J. Rock Mech. Min. Sci. 34 (8) (1997) 1165–1186, http://dx.doi.org/10.1016/S1365-1609(97)80069-X.
[24] J. Spross, T. Gasch, F. Johansson, Implementation of reliability-based thresholds to excavation of shotcrete-supported rock tunnels, Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards 17 (2) (2022) 361–375, http://dx.doi.org/10.1080/17499518.2022.2046789.
[25] K. Esmaieli, J. Hadjigeorgiou, M. Grenon, Estimating geometrical and mechanical REV based on synthetic rock mass models at brunswick mine, Int. J. Rock Mech. Min. Sci. 47 (6) (2010) 915–926, http://dx.doi.org/10.1016/j.ijrmms.2010.05.010.
[26] C. Edelbro, F. Perman, B. Figueiredo, J. Sjöberg, Evaluation and Interpretation of Initial Rock Stresses for Stockholm and Gothenburg, BeFo Report 231, 2022.
[27] A. Palmström, H. Stille, Rock Engineering, Thomas Telford Ltd, London, 2010.
[28] FLAC3D — Fast Lagrangian Analysis of Continua in Three-Dimensions, Itasca Consulting Group, Inc., Minneapolis, 2019, Ver. 7.0.
[29] W. Zhang, A.T. Goh, Reliability assessment on ultimate and serviceability limit states and determination of critical factor of safety for underground rock caverns, Tunn. Undergr. Space Technol. 32 (2012) 221–230, http://dx.doi.org/10.1016/j.tust.2012.07.002.
[30] J.C. Langford, M. Diederichs, Reliability based approach to tunnel lining design using a modified point estimate method, Int. J. Rock Mech. Min. Sci. 60 (2013) 263–276, http://dx.doi.org/10.1016/j.ijrmms.2012.12.034.
[31] Q. Lü, C.L. Chan, B.K. Low, System reliability assessment for a rock tunnel with multiple failure modes, Rock Mech. Rock Eng. 46 (2013) 821–833, http://dx.doi.org/10.1007/s00603-012-0285-3.
[32] W. Bjureland, J. Spross, F. Johansson, A. Prästings, S. Larsson, Reliability aspects of rock tunnel design with the observational method, Int. J. Rock Mech. Min. Sci. 98 (2017) 102–110, http://dx.doi.org/10.1016/j.ijrmms.2017.07.004.
[33] Y. Fang, Y. Su, Y. Su, S. Li, A direct reliability-based design method for tunnel support using the performance measure approach with line search, Comput. Geotech. 107 (2019) 89–96, http://dx.doi.org/10.1016/j.compgeo.2018.11.018.
[34] E.M. Dawson, W.H. Roth, A. Drescher, Slope stability analysis by strength reduction, Géotechnique 49 (6) (1999) 835–840, http://dx.doi.org/10.1680/geot.1999.49.6.835.

BeFo

Box 5501
SE-114 85 Stockholm

info@befoonline.org • www.befoonline.org
Besöksadress: Storgatan 19, Stockholm

ISSN 1104-1773