KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

TAIKOMOSIOS MATEMATIKOS KATEDRA

Vaida Česnulytė

EXTREME PRECIPITATION PROBABILISTIC ASSESSMENT AND UNCERTAINTY ANALYSIS

Magistro darbas

Vadovas
dr. R. Alzbutas

KAUNAS, 2011

KAUNO TECHNOLOGIJOS UNIVERSITETAS

FUNDAMENTALIŲJŲ MOKSLŲ FAKULTETAS

TAIKOMOSIOS MATEMATIKOS KATEDRA

TVIRTINU
Katedros vedėjas
doc. dr. N. Listopadskis
2011 06 02

EXTREME PRECIPITATION PROBABILISTIC ASSESSMENT AND UNCERTAINTY ANALYSIS

Taikomosios matematikos magistro baigiamasis darbas

Vadovas
dr. R. Alzbutas
2011 06 01

Recenzentas
doc.dr. R. Krikštolaitis
2011 06 01

Atliko
FMMM-9 gr. stud.
V. Česnulytė
2011 05 30

KAUNAS, 2011

KVALIFIKACINĖ KOMISIJA

Pirmininkas: Leonas Saulis, profesorius (VGTU)

Sekretorius: Eimutis Valakevičius, docentas (KTU)

Nariai: Algimantas Jonas Aksomaitis, profesorius (KTU)

Vytautas Janilionis, docentas (KTU)

Vidmantas Povilas Pekarskas, profesorius (KTU)

Rimantas Rudzkis, habil. dr., vyriausiasis analitikas (DnB NORD Bankas)

Zenonas Navickas, profesorius (KTU)

Arūnas Barauskas, dr., vice-prezidentas projektams (UAB „Baltic Amadeus“)

Magistro darbo išplėstinė santrauka

Prasidėjus XXI amžiui vis didesnį visuomenės susidomėjimą kelia klimato pokyčiai. Šiandien mokslininkai didžiulį dėmesį skiria gamtos tyrinėjimams, ypač po to, kai padaugėjo stichinių ir katastrofinių gamtos nelaimių. Viena iš skaudžiausių pastarųjų gamtos katastrofų buvo šių metų kovo 11-osios dienos žemės drebėjimas Japonijoje, pasiekęs 9.0 balo pagal Richterio skalę [28]. Japinijos užsienio reikalų ministerijos (Ministry of Foreign Affairs in Japan) duomenimis šios katastrofos metu žuvo 15 093 žmonės [28]. Ši ir kitos gamtos katastrofos (Eijafjalajokutlio ugnikalnio išsiveržimas Islandijoje 2010 03 21 [7], cunamis Indijos vandenyne 2004 12 26 [4], miškų gaisrai Australijoje 2006-2007 m. [3] ir kt.) parodo, kokie svarbūs yra klimato tyrinėjimai, jų prognozavimas ir išankstinis galimų pasekmių ir žalos įvertinimas.

Gamtos elgesį bandoma nuspėti įvairiausiais statistiniais, matematiniais ir eksperimentiniais metodais. Yra kuriama daugybė modelių, kurių pagalba galima apskaičiuoti įvairias gamtos elgesio charakteristikas. Deja, nepakanka sukurti modelį, apskaičiuoti norimą charakteristiką ir padaryti išvadas. Dėl daugybės trikdžių, pašalinių veiksnių įtakos, skaičiavimo ir matavimo netikslumų gauti rezultatai ne visada atitinka tikrovę.

Pagrindinis **šio darbo tikslas** yra Dūkšto regione pasitaikančios ekstremalios sniego dangos tikimybinis vertinimas bei neapibrėžtumo ir jautrumo analizė. Dūkšto regionas pasirinktas dėl duomenų gausos ir analizės aktualumo, kuri yra susijusi su Ignalinos atominės elektrinės (IAE) ir būsimos Visagino atominės elektrinės (VAE) aikštelių rizikos vertinimu. Be IAE Lietuvoje yra nemažai didelio saugumo reikalaujančių statinių: Kauno hidroelektrinė, Kruonio hidroakumuliacinė elektrinė, „Mažeikių naftos“ terminalas, „Achemos“ pastatų grupė ir kt. Dėl to būtina atsižvelgti į gausios sniego dangos atsiradimo tikimybę.

Bendradarbiaujant su Lietuvos Hidrometeorologijos Tarnyba (LHMT) ir Dūkšto meteorologijos stotimi (Dūkšto MS) buvo naudojami Dūkšto ekstremalios sniego dangos duomenys (1992-2009 m.), todėl modeliavimo ir analizės rezultatai kur kas geriau atitinka dabartines sąlygas.

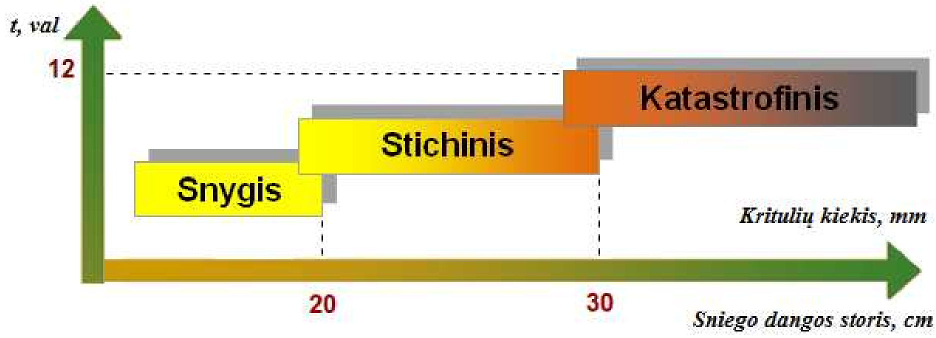


Fig. 1. Sniego intensyvumo kriterijus

Ekstremalus įvykis – tai nustatytus kriterijus, pasiekęs ar viršijęs gamtinio, techninio, ekologinio ar socialinio pobūdžio įvykis, keliantis pavojų žmonėms, jų fiziologinėms ar socialinėms gyvenimo sąlygoms, turtui, ūkiui ir aplinkai [42]. Ekstremalūs įvykiai skirstomi į stichinius ir katastrofinius. Pirmame paveiksle (Fig. 1) pavaizduota, kokiam sniego intensyvumui esant jis tampa stichiniu arba katastrofiniu. Šiame darbe nagrinėjami ekstremalūs sniego svoriai atitinka maksimalias metines sniego svorio reikšmes.

Ekstremaliam sniego svoriui apibūdinti yra naudojami apibendrintieji ekstremalių reikšmių skirstiniai, kurie yra trijų pagrindinių tipų: I, II ir III (Appendix 3). I-ojo tipo Gumbelio skirstinys naudojamas maksimalaus sniego svorio reikšmių tikimybėms apskaičiuoti per ilgesnį laikotarpį. Gumbelio skirstinys yra atskiras Fišerio-Tipeto (Fisher-Tippett) skirstinio, dar kitaip vadinamo log-Veibulo (log-Weibull) skirstiniu, atvejis. II-ojo tipo Frečeto (Fréchet) skirstinys yra aprėžtas iš apačios, be sniego svorio dar naudojamas formos (shape) parametras, gali būti pridėti skalės (scale) ir išsidėstymo (location) parametrai. III-iojo tipo Veibulo skirstinys sniego svorio tyrime naudojamas plačiausiai, nes įvairios jo modifikacijos geriausiai atitinka tam tikrą vietą ir laiką, be to puikiai tinka vertinti metinį sniego svorį. Veibulo skirstinys yra aprėžtas iš viršaus ir gali turėti du arba tris parametrus. Visų šių trijų tipų skirstiniai yra lengvai pakeičiami vienas kitu, panaudojant normalųjį logaritmą. Be šių trijų pagrindinių tipų skirstinių yra naudojama daugybė kitų, specialiai pritaikytų prie vietos sąlygų, klimatinės situacijos ir kt. Tačiau dauguma jų yra paremti ekstremalių reikšmių skirstiniais.

Sudarant ir taikant tikimybinį modelį naudojama įvairi informacija. Vienas iš pagrindinių informacijos šaltinių yra išmatuoti nagrinėjamo įvykio duomenys: sniego dangos storis, tankis, vandens atsargos sniege. Matuojant šiuos duomenis tiesiogiai yra neišvengiama ne tik matavimo prietaisų paklaidų, dar vadinamų sistemingosiomis, bet ir atsitiktinių paklaidų, kurių priežastys yra atsitiktiniai ir tyrimo metu nekontroliuojami trikdžiai.

Sistemines paklaidas sąlygoja prietaisų netikslumai, eksperimento metodikos netobulumai, patikros ir reguliavimo netikslumai, teorinio modelio neatitikimas aprašomajam fizikiniam procesui ir kt. Atsitiktinės paklaidos yra susijusios su vėjo stiprumu, geografine padėtimi (sniego aukštis atvirose vietose matuojamas kas 20 metrų, miške – kas 10 metrų [27, 13.6.8 punktas]), laikotarpiu (sniego nuotraukos žiemą daromos kas 10 dienų, tirpimo periodu – kas 5 dienas [27, 13.6.4 punktas]). Dėl šių priežasčių gautas rezultatas nėra 100% tikslus. Todėl yra būtina neapibrėžtumo analizė. Be to, rezultato netikslumui įtakos turi ir nepakankamas duomenų kiekis, skaičiavimų klaidos, klaidingos išvados ir kt. Lietuvoje neapibrėžtumo analizė dar nėra taip plačiai naudojama kaip užsienyje. Kitose pasaulio šalyse bene kiekvieno tyrimo, projekto ar analizės darbus lydi neapibrėžtumo vertinimas. Ypač tai yra svarbu atliekant tikimybinio modelio taikymą, kai reikia įvertinti ne tik modeliavimo, bet ir parametrų neapibrėžtumą. Vieni iš dažniausiai naudojamų metodų neapibrėžtumo analizei atlikti yra šie: diferencialinės analizės metodas, Gryno (Green‘s) funkcijos metodas, tiesioginio sujungimo-išskyrimo metodas, Monte Karlo (Monte Carlo) atrankos metodas, kompiuterinės algebros metodas ir kt. Šiam darbui atlikti yra sukurta nemažai programinės įrangos: ADIFOR, AIRDOS, COSYMA, GRESS, SUSA, SIMLAB, WASP ir kt.

Tam, kad būtų galima sumodeliuoti gamtos procesus, reikia turėti modelius, kurie atspindėtų realių elementų esmines savybes ir rezultatai atspindėtų realioje erdvėje vykstančius procesus, tačiau pats matematinis modelis būtų kuo paprastesnis. Įvykių modeliavimas yra skirstomas į fizikinį ir statistinį. Fizikinis modeliavimas yra paremtas matematinių lygčių, aprašančių nagrinėjamus procesus, sprendimas. Statistinis modeliavimas yra toks, kurio metu analizuojami ilgalaikiai įvykių charakteristikų matavimų duomenys, nustatomi statistiniai ryšiai tarp prognozės rezultatų ir faktinių duomenų. Matematinius modelius dažnai patogu užrašyti kaip funkciją [19]:

$$y = F(x_1, x_2, \ldots, x_N), \tag{0.1}$$

kur

$x_1, x_2, \ldots, x_N$ – modelio parametrai;

N – modelio parametrų skaičius;

y – modelio rezultatas;

$F(\cdot)$ – funkcija, siejanti modelio parametrus ir modelio rezultatą.

Matematinis modelis (0.1) yra šiek tiek supaprastintas, nes realiuose modeliuose dažniausiai yra ne vienas rezultatas, o daug rezultatų, apibūdinančių įvairias nagrinėjamo proceso ar reiškinio charakteristikas, t. y. y yra vektorius, o ne skaliarinis dydis. Taip pat dažnai reikia analizuoti rezultatus, kurie priklauso nuo laiko t, t. y. $y=y(t)$, o funkcija F tuomet taip pat priklauso nuo t. Tačiau supaprastintas modelis (0.1) nekeičia neapibrėžtumo ir jautrumo analizės principų.

Ekstremalaus sniego svorio tikimybinėje analizėje pagrindinė užduotis yra apžvelgti įvairių sniego svorių pasirodymo tikimybes. Šiuo atveju sniego svorio x pasirodymo tikimybė yra apibrėžiama ekstremalių reikšmių skirstiniais $F(x)$ (2.1.1, 2.1.2, 2.1.3, 2.1.4 skyreliai):

- Gumbelio

$$F(x) = P(X < x) = e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}, \ -\infty < x < \infty, \sigma > 0; \tag{0.2}$$

- Veibulo

$$F(x) = P(X < x) = 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^{\beta}}, \ x \geq \mu, \sigma > 0, \beta > 0; \tag{0.3}$$

- Frečeto

$$F\left(x\right)=P\left(X<x\right)=e^{-\left(\frac{\sigma}{x-\mu}\right)^{\beta}},\ x\geq\mu,\sigma>0,\beta>0; \tag{0.4}$$

- Apibendrintojo ekstremalių reikšmių

$$F\left(x\right)=P\left(X<x\right)=\exp\left\{-\left[1+\frac{\beta\left(x-\mu\right)}{\sigma}\right]^{-\frac{1}{\beta}}\right\}, \tag{0.5}$$

$$1+\frac{\beta\left(x-\mu\right)}{\sigma}>0,\quad \sigma>0, \tag{0.6}$$

kur μ yra išsidėstymo (location), $\sigma>0$ skalės (scale) ir $\beta>0$ formos (shape) parametrai.

Modelio parametrai yra apibrėžiami statistiškai ir pasiskirstymo dėsnis pasirenkamas atitinkamiems statistiniams kriterijams.

Nagrinėjant ekstremalias reikšmes naudojamas skirstinio $F(x)$ papildinys – garantijų funkcija $G(x)$, kuri parodo tikimybę, kad sniego svoris bus didesnis už x:

$$G\left(x\right)=P\left(X>x\right)=1-P(X<x)=1-F(x). \tag{0.7}$$

Dydis atvirkščias garantijų funkcijai $G(x)$ vadinamas pasikartojimo (grįžimo) periodu $E\left(T\right)$ (2.1.5 skyrelis):

$$E\left(T\right)=\frac{1}{1-F(x)}=\frac{1}{G\left(x\right)}. \tag{0.8}$$

Pasikartojimo periodas $E\left(T\right)$ parodo laiko tarpą, per kurį su apskaičiuota tikimybe yra tikėtinas ekstremalaus sniego svorio pasirodymas. Žinant sąryšį tarp garantijų funkcijos $G\left(x\right)$ ir pasikartojimo periodo $E\left(T\right)$ yra apskaičiuojamos galimos ekstremalaus sniego svorio reikšmės x skirtingais pasikartojimų periodais:

$$x_{Gumbelio}=\mu+\sigma\left(-\ln\left(\ln\left(E(T)\right)-\ln\left(E(T)-1\right)\right)\right) \tag{0.9}$$

$$x_{Veibulo}=\mu+\sigma(\ln\left(E(T)\right)-\ln\left(E(T)-1\right))^{1/\beta}, \tag{0.10}$$

$$x_{Frečeto}=\mu+\frac{\sigma}{\left(\ln\left(E(T)\right)-\ln\left(E(T)-1\right)\right)^{\beta}}, \tag{0.11}$$

$$x_{Apibendrintojo}=\mu+\frac{\sigma}{\beta(\ln\left(E(T)\right)-\ln\left(E(T)-1\right))^{\beta}}-\frac{\sigma}{\beta}, \tag{0.12}$$

kur $E\left(T\right)$ yra pasikartojimo periodas.

Per 100 metų galimas (ir didesnis) ekstremalus sniego svoris yra 280,9 kg/m^2 (Gumbelio), 263,4 kg/m^2 (Veibulo), 321 kg/m^2 (Frečeto) ir 320,5 kg/m^2 (Apibendrintojo).

Siekant įvertinti skristinio tinkamumą turimiems duomenims yra naudojami įvairūs metodai:

- Grafiniai metodai (Fig. 6):
  - *P-P diagrama* (empirinės ir teorinės pasiskirstymo funkcijos palyginimo grafikas);
  - *Q-Q diagrama* (empirinių ir teorinių kvantilių palyginimo grafikas);
- Suderinamumo hipotezių tikrinimas [1]:
  - *Chi kvadratu suderinamumo kriterijus*;
  - *Kolmogorovo-Smirnovo suderinamumo kriterijus;*
  - *Andersono-Darlingo suderinamumo kriterijus.*

P-P ir Q-Q diagramos parodo tai, kad kuo arčiau tiesės išsidėsto taškai, tuo geresnis empirinio skirstinio $F_n(x)$ suderinamumas su teoriniu skirstiniu $F(x)$.

Atsižvelgiant į suderinamumo statistikas ir palyginimą tarp empirinio ir teorinio skirstinių, gauta, jog optimalus skirstinys vertinant ekstremalaus sniego svorio reikšmes yra **Frečeto** skirstinys.

Norint vykdyti neapibrėžtumo analizę, reikalinga identifikuoti ir apibūdinti neapibrėžtumo šaltinius. Šiam tikslui palengvinti siūloma atsižvelgti į neapibrėžtumų klasifikavimą pagal jų atsiradimo priežastis ir pasekmes. Neapibrėžtumų kilmė gali būti siejama su sistemos modelio parametrų matavimo paklaidomis bei informacijos trūkumu arba natūralia, bet nevaldoma parametrų variacija. Nagrinėjant sistemos modelio analizės rezultatų neapibrėžtumą, siūloma atsižvelgti į visą rezultatų atsiradimą sąlygojančią seką (Fig. 2), kurią sudaro sistema, modeliavimo procesas, modelis ir modelio analizė [2].

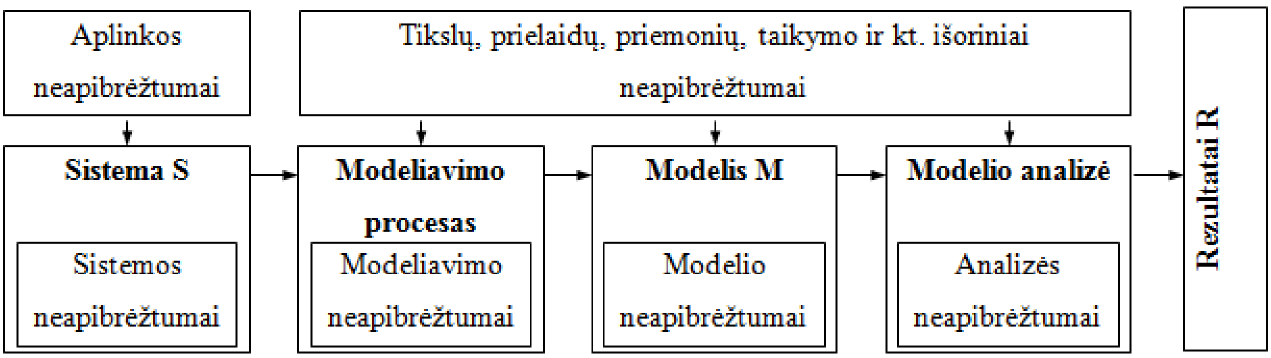


Fig. 2. Sistemos modelio rezultatų neapibrėžtumą sąlygojanti seka

Atsižvelgus į neapibrėžtumo valdymo galimybes, išskiriami išoriniai ir vidiniai neapibrėžtumo šaltiniai bei atitinkami modelio parametrai. Kuo mažesnės neapibrėžtumo šaltinių (kraštinių sąlygų, parametrų ir kt.) valdymo galimybės, tuo labiau jie laikomi išoriniais.

Kadangi ne visi neapibrėžtumai yra įvertinami tikimybiniais skirstiniais, todėl iš anksto turi būti numatyta, kaip atsižvelgti į tokią situaciją. Tam gali būti panaudotos mažiausiai dvi alternatyvos. Pirmoji iš jų grindžiama naudojimu standartinių, jau priimtų metodų, kuriuose naudojamos prielaidos ir modeliai yra laikomi priimtinais. Antroji alternatyva gali leisti modelių ir prielaidų įvairovę, bet reikalauti, kad būtų atliekama jautrumo analizė siekiant nustatyti, kaip pasikeistų rezultatai ir jų pagrindu daromos išvados. Tokiu būdu galutinis sprendimas priimamas įvertinus skirtingų modelių ir prielaidų santykinę įtaką į daromų išvadų skirtumus. Taigi potencialios neapibrėžtumų pasekmės siejamos su rezultatų jautrumo analize bei rezultatams ir modelio parametrams taikomais priimtinumo kriterijais.

Kiekybiškai neapibrėžtumo analizės rezultatai gali būti išreikšti gautojo skirstinio kvantiliais (pvz., 5% ir 95%), kurie yra nustatomi žinant konkretų skirtinį. Praktiškai modelio rezultatų pasiskirstymo funkcija gali būti gaunama naudojant parametrų subjektyvius tikimybinius skirstinius ir Monte Karlo tipo metodus.

Norint įvertinti galimą modeliavimo klaidų poveikį, įprastai yra apskaičiuojamos (α, β) statistinės tolerancijos ribos. Čia β yra pasikliautinumo lygmuo, kurio maksimalus modelio rezultatas neviršys su tikimybe α (arba α % kvantiliu, kiekybiškai apibūdinančiu bendrą visų nustatytų neapibrėžtumų įtaką). Pavyzdžiui, pagal Vilkso formulę [37], pakanka 93 eksperimentų, kad būtų gaunamas (0,95; 0,95) statistinės tolerancijos intervalas. Šiame darbe yra atliekama 100 eksperimentų minėtam tolerancijos intervalui gauti. Bendru atveju eksperimentų skaičius n_1, kuris naudojamas norint gauti vienpusės tolerancijos ribas ir eksperimentų skaičius n_2, kuris naudojamas norint gauti dvipusės tolerancijos intervalus, gali būti išreikšti pagal Vilkso formulę taip:

$$n_1 \geq \frac{\ln(1-\beta)}{\ln(\alpha)}, \tag{0.13}$$

$$n_2 \geq \left(\ln(1-\beta) - \ln\left(\left(\frac{n_2}{\alpha} \right) + 1 - n_2 \right) \right) / \ln(\alpha). \tag{0.14}$$

Mažiausias eksperimentų skaičius (Table 1), reikalingas šioms riboms nustatyti, nepriklauso nuo neapibrėžtų nežinomųjų ir priklauso tik nuo tikimybių α ir β, pateiktų aukščiau. Eksperimentų skaičius yra neparametrinės statistikos rezultatas. Privalumas yra tas, kad šis skaičius visiškai nepriklauso nuo neapibrėžčių kiekio ir neturi įtakos pagrindiniam skirstiniui.

Matematinio modelio jautrumo analizė skirta tirti labiausiai modelio rezultato neapibrėžtumą sąlygojančius veiksnius. Pagal jautrumo analizės rezultatus galima nustatyti, kurių modelio parametrų tikslesnis įvertinimas leistų ženkliai sumažinti modelio rezultato neapibrėžtumą ir kurių parametrų tolesnis tikslinimas nėra prasmingas dėl jų mažos įtakos rezultatui. Kadangi parametrų įvertinimas susijęs su turimomis žiniomis apie reiškinius ar fizikinius dydžius, tai parametrui tiksliau įvertinti gali tekti atlikti papildomus eksperimentinius tyrimus.

Jautrumo analizės metodai yra dviejų tipų: lokalūs ir globalūs. Pastarieji skirstomi į imties ir dispersijos išskaidymo [20]. Vienas populiariausių imties metodų jautrumo analizėje yra standartizuota tiesinė regresija. Matematinį modelį:

$$1-\alpha^{n}-n(1-\alpha)\alpha^{n-1}\geq\beta \tag{0.15}$$

išreiškus daugialype tiesine parametrų funkcija gaumane:

$$y=F\left(x_1,x_2,\ldots,x_N\right)=\alpha+b_1x_1+\ldots+b_Nx_N. \tag{0.16}$$

Koeficientai b_i apskaičiuojami mažiausių kvadratų metodu (MKM), tačiau jie negali būti jautrumo indeksais, nes būtina normuoti parametrų matavimo skales. Parametrų matavimo vienetai normuojami standartizuojant kiekvieną parametrą ir modelio rezultatą [20]:

$$\widehat{x}_{i,k}=\frac{x_{i,k}-Ex_i}{\sigma x_i}, i=1,2,\ldots,N; k=1,2,\ldots,M; \tag{0.17}$$

$$\widehat{y}_k=\frac{y_k-Ey}{\sigma y}; \tag{0.18}$$

čia

$Ex_i - x_i$ yra parametro vidurkis;

Ey – modelio rezultato vidurkis;

σx – parametro standartinis nuokrypis;

σy – modelio rezultato standartinis nuokrypis;

M – parametrų atsitiktinės imties dydis;

N – parametrų skaičius.

Tuomet regresijos koeficientai β_i standartizuotiems dydžiams vadinami standartizuotais regresijos kodeficientais (SRK) ir yra dažnai naudojami parametrų jautrumo indeksais:

$$\widehat{y}=\alpha+b_1\widehat{x}_1+\ldots+b_N\widehat{x}_N. \tag{0.19}$$

Standartizuoti regresijos (SRK) ir dalinės koreliacijos koeficientų (DKK) rezultatai yra paremti modelio tiesiškumo hipoteze. Norint patvirtinti šią hipotezę yra svarbu apskaičiuoti tiesinio modelio koeficientus:

$$R^2=\frac{\sum_{i=1}^{N}\left(\widehat{y}_i-\overline{y}\right)^2}{\sum_{i=1}^{p}\left(y_i-\overline{y}\right)^2}, \tag{0.20}$$

kur $\widehat{y}_i$ yra y_i įvertis iš regresijos modelio.

Determinacijos koeficientas R^2 parodo, kaip tiksliai tiesinės regresijos modelis atspindi tikrojo Y gavimą. Kuo R^2 yra arčiau 1, tuo modelio tikslumas yra didesnis. Praktikoje dažnai reikalaujama, kad tiesinio modelio determinacijos koeficientas būtų ne mažesnis nei 0,6, t. y. parametrų neapibrėžtumas paaiškintų ne mažiau nei 60% modelio rezultato neapibrėžtumo. Esant mažam R^2, koreliacijos koeficientai taip pat negali būti parametrų jautrumo indeksais.

Remiantis Vilkso formule ir taikant neapibrėžtumo ir jautrumo analizės metodiką [13], kiekvienam sniego svoriui modeliavimų kiekis buvo parinktas toks, kad rezultato neapibrėžtumą būtų galima įvertinti taikant 0,95 pasikliovimo lygmenį. Taikant aprašytą metodiką [13], kiekvienam sniego svoriui sumodeliuota 100 skirtingų galimų variantų, turint 100 skirtingų parametrų porų, t.y. kiekvienam sniego svoriui skaičiuojant Frečeto tikimybės įvertį naudota 100 skirtingų μ, σ ir β parametrų, kurie apskaičiuoti pagal 8 lentelėje (Table 8) nurodytus tikimybinius skirstinius. Dėl duomenų apie ekstremalius sniego svorius trūkumo, kiekvieniems metams buvo sugeneruota po 100 ekstremalių reikšmių pagal Frečeto skirstinį su parametrais, apskaičiuotais 2 lentelėje (Table 2). Iš viso sumodeliuoti 8 sniego svorio lygiai su skirtingomis tam lygiui sniego svorio reikšmėmis x: 1, 10, 50, ..., 300 kg/m^2. Modelio rezultatai vaizduojami tikimybių kitimo grafike, kuriame pateikiami su 100 skirtingų parametrų porų kiekvienam sniego svoriui gauti įverčiai (Fig. 14). Logaritminėje skalėje pavaizduotos 100 kreivių vizualiai parodo, koks gali būti neapibrėžtumas.

Vertinant rezultatų jautrumą pradinių parametrų neapibrėžtumui buvo naudojami regresijos ir koreliacijos koeficientai: standartizuotas regresijos koeficientas (SRK) ir dalinės koreliacijos koeficientas (DKK). Tiek SRK, tiek DKK parodė, kad nagrinėjant ekstremalų sniego svorį didžiausią įtaką modelio rezultatui turi tas modelio parametras, kuris atitinka σ parametrą (Fig. 16, Fig. 17).

Vykdant neapibrėžtumo ir jautrumo analizę vienas svarbiausių koeficientų yra determinacijos koeficientas R^2, kuris parodo, kurią modelio rezultatų sklaidos dalį galima modeliuoti taikant tiesinę regresiją. Apskaičiavus determinacijos koeficientą R^2 nustatyta, kad, kai sniego svoris ne didesnis nei 110 kg/m^2, tai iki 90 procentų tikimybinio modelio neapibrėžtumo gali būti nustatyta taikant regresinį modelį. Didėjant sniego svoriui determinacijos koeficientas R^2 mažėja, o tai rodo, kad regresinis modelis vis silpniau nusako parametrų ir rezultato neapibrėžtumo sąryšį. Taip yra dėl didelės parametrų sklaidos ir mažo informacijos kiekio apie tokius įvykius.

Darbo išvados

1. Per 100 metų laikotarpį galimas (ir didesnis) sniego dangos svoris yra: 280,9 kg/m^2 (Gumbelio), 263,4 kg/m^2 (Veibulo), 321 kg/m^2 (Frečeto) ir 320,5 kg/m^2 (Apibendrintojo).
2. 50 kg/m^2 sniego svorio pasikartojimo periodas yra 2 metai. 100 kg/m^2 sniego svorio pasikartojimo periodas yra: 3 metai (Gumbelio) ir 4 metai (Veibulo, Frečeto, Apibendrintojo).
3. Pritaikius empirinio ir teorinio skirstinių palyginimą bei įvertinus hipotezių suderinamumo kriterijus pastebėta, kad optimalus skirstinys vertinant sniego svorį yra Frečeto skirstinys.
4. Atlikta jautrumo analizė rodo, kad ekstremalaus sniego tikimybės įverčio vertes ir jų neapibrėžtumą visais atvejais labiausiai įtakoja skalės σ parametras (Par. 2).
5. Apskaičiavus determinacijos koeficientą nustatyta, kad kai sniego svoris ne didesnis nei 110 kg/m^2, tai iki 90 procentų tikimybinio modelio neapibrėžtumo gali būti nustatyta taikant regresinį modelį. Didėjant sniego svoriui determinacijos koeficientas mažėja, o tai rodo, kad regresinis modelis vis silpniau nusako parametrų ir rezultato neapibrėžtumo sąryšį.
6. Atliktas darbas gali būti panaudotas vertinant tokią pavojingo objekto rizikos analizę, kuomet atsižvelgiama ne tik į vidinius, bet ir į tam tikroje teritorijoje pasireiškiančius išorinius įvykius. Tam, visų pirma, reikia išanalizuoti šiai teritorijai būdingus duomenis ir perskaičiuoti pavojingo kritulių kiekio tikimybes bei įvertinti pasekmes ir modelio rezultatų neapibrėžtumą pagal šiame darbe aprašytą metodiką.

Table of Contents

List of Tables

List of Figures

List of publications

1. **V.Česnulytė**, R.Alzbutas. *Extreme precipitation probability assessment and uncertainty sensitivity analysis.* CYSENI 2010. The 7th Annual Conference of Young Scientists on Energy Issues, May 27-28, 2010 Lithuanian Energy Institute, Kaunas, Lithuania;

2. **V.Česnulytė**, R.Alzbutas. *Probability assessment and uncertainty and sensitivity analysis of extreme rainfall.* Student Research Works. 2008-20410. Conference proceedings. ISBN 978-609-420-096-4.

3. **V.Česnulytė**, R.Alzbutas. *Extreme rainfall probability assessment*. Applied Mathematics. VIII student research conference proceedings. ISBN 978-9955-25-811-7.

4. **V.Česnulytė**, R.Alzbutas. *Extreme value distributions for probabilistic analysis of extreme snowfall.* CYSENI 2011. The 8th Annual Conference of Young Scientists on Energy Issues, May 26-27, 2011 Lithuanian Energy Institute, Kaunas, Lithuania.

Acronyms

CDF	Cumulative Distribution Function
df	Degrees of freedom
ECDF	Empirical Cumulative Distribution Function
EDF	Empirical Distribution Function
EVT	Extreme Value Theory
GEV	Generalized Extreme Value Distribution
GOF	Goodness of fit
MS	Meteorological Station
NPP	Nuclear Power Plant

Introduction

Modern man is making all efforts for the maximum benefit and for feeling safe. Unfortunately, there are certain factors that can not be controlled, but they influence the life. Nature is one of such factors. Today scientists have a huge emphasis on exploration of nature, especially after the rise of natural and catastrophic natural disasters. The attempts to predict the behavior of nature are made by a wide range of statistical, mathematical, and experimental methods. A number of models that can help calculating a range of natural behavioral characteristics are being developed. Unfortunately, it is not enough to create a model, to calculate desired characteristics, and to draw the conclusions. The obtained result does not always review the reality because of a variety of interference of extraneous factors, calculating and measuring errors.

Extremes in nature occur almost worldwide and cause incalculable human losses, in addition to billions of dollars damages each year (e.g. 622,000 people died from natural disasters in ten years period from 1992) [38]. The effects of other weather-related hazards, such as: windstorms, hurricanes, typhoons, and landslides are also known to be severe. Extreme events loss from acts of nature is increasing fast. There has been a rising trend in the number of events and economic and insurance losses from 1960 to 1990 [38].

The **main aim** of this paper is to investigate the probabilistic model of extreme weight of snow according to Extreme Value distributions and to find out which Gumbel's, Weibull's, GEV's, and Fréchet's distribution the best fits for the data. The probabilities of weight of snow, posing a serious environmental hazard are estimated there. The extreme snow thickness data at Dūkštas region (1992-2009) were measured in cooperation with Lithuanian Hydrometeorological Service and Dūkštas meteorological station (MS); so the results of the modeling and analysis better suit for current conditions.

The choice of this natural disaster was because of the high risk of accidents caused to such essential objects as Ignalina Nuclear Power Plant, Kaunas Hydroelectric Power Plant, Kruonis Pump Storage Power Plant, "Mažeikiai nafta" terminal, "Achema" group of buildings, etc. No matter the importance and significance of this natural phenomenon, much less attention is paid to them than to the other elements.

Salvadori (2007) [38] mentioned that the study of the statistics of extreme events is the first step in the mitigation of these national disasters. In addition to historical records containing observations from the past it is usually the only source of information. It is expected to develop and to apply statistical technique in order to estimate the degree of risk involved globally using limited samples of the data and imperfect knowlegde of the involved processes.

1. ANALYTICAL PART

1.1. Extreme Value Theory

Extreme value theory is a branch of statistics dealing with the extreme deviations from the median of probability distributions. The general theory sets out to assess generated by processes type of probability distributions. Extreme value theory is important for assessing risk for highly unusual events (e. g. 100-year floods).

The key to Extreme Value Theory (EVT) is the extreme value theorem (a cousin of the better-known central limit theorem) which tells us about the distribution limit of extreme value as our sample size increases. Suppose we have some return observations but do not know the density function from which they are drawn. Subject to the certain relatively innocuous conditions, this theorem tells us that the distribution of extreme returns converge asymptotically to [10]:

$$F_{\beta,\mu,\sigma}(x)=\begin{cases}\exp\left(-\left[1+\dfrac{\beta(x-\mu)}{\sigma}\right]^{-\frac{1}{\beta}}\right) if\,\beta\neq 0\\ \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right) if\,\beta=0\end{cases}. \tag{1.1}$$

The parameters μ and σ correspond to the mean and standard deviation, and the third parameter β gives an indication of the heaviness of the tails: the bigger β, the heavier the tail. In this case, our asymptotic distribution takes the form of Fréchet distribution.

This theorem tells us that the limiting distribution of extreme returns always has the same form – whatever the distribution of the parent returns where our extreme returns are drawn. It is important because it allows to estimate extreme probabilities and extreme quantiles without having to make effective assumptions about the unknown parent distribution.

EVT deals with the frequency and magnitude of a very low probability events. Precisely as extreme events are extreme, we have to operate with a very small data sets, and this means that our estimates – our quantile estimates and the estimated probabilities associated with them – are inevitably very imprecise. However, EVT makes the best out of an inherently difficult problems.

1.2. The importance of extreme values

Natural calamities of great magnitude happen all over the world (e.g. the extraordinary dry spell in the western regions of the United States and Canada during the summer of 2003, the devastating earthquake that destroyed almost the entire historic Iranian city of Bam in 2003 [17], the massive snow fall in the eastern regions of the United States and Canada during February 2004 [9]) or the destructive hurricanes and devastating floods affecting many parts of the world. For this reason, an architect may be interested in constructing a high rise building withstanding an earthquake of great magnitude, maybe a ‘‘100-year earthquake’’; or an engineer building a bridge across the Mississippi river may be interested in fixing its height so that the water may be expected to go over the bridge once in 200 years. It is evident that the characteristics of interest in all these cases are the extremes corresponding to either minimum or maximum values.

Knowledge of extreme relevant phenomena distributions is important in obtaining good solutions to engineering design problems. However, estimating extreme capacities or operating conditions is very difficult because of the lack of available data. Note that engineering design must be based on extremes, because the largest values (loads, earthquakes, winds, floods, waves, etc.) and the smallest values (strength, supply, etc.) are the key parameters leading to the failure of engineering works.

Castillo and Gumbel noticed that there are many areas where extreme value theory plays an important role [6], [12]:

- ***Ocean engineering***. The design of offshore platforms, breakwaters, dikes, and other harbor works rely upon the knowledge of the probability distribution of the highest waves. The joint distribution of the heights and periods of the seawaves is the other problem of crucial interest .
- ***Structural engineering***. A correct structure design requires a precise estimation of extreme winds occurrence probabilities, loads or earthquakes and seismic incidence, consequences of extreme weight of snow in order to allow realistic safety margins on one hand and for economical solutions on the other.
- ***Hydraulics engineering***. Quantifying uncertainty in flood magnitude estimators is an important problem in floodplain development, including risk assessment for floodplain management, risk-based design of hydraulic structures and estimation of expected annual flood damages.
- ***Material strength***. The analysis of size effect is one interesting application of extreme value theory to material strength. The strength of actual structures has to be inferred in many engineering problems from the strength of small elements of reduced size samples, prototipe or models, which are tested under the laboratory conditions. Extrapolation from small to much larger sizes is needed. In this context, the extreme value theory becomes very useful in order to analize the size effect and to make extrapolations not only possible but also reliable.

- ***Fatigue strength***. Modern fracture mechanics theory reveals that fatigue failure becomes because of cracks propagation when the elements are under the action of repetitive loads. The fatigue strength of a piece is governed by the largest crack in the piece. The presence of cracks in pieces is random in number, size, and shape, and, thus, resulting in a random character or fatigue strength.
- ***Electrical strength of materials***. Lifetime of some electrical devices depends not only on their random quality, but also on the random voltage levels acting on them. The device survives if the maximum voltage level does not surpass critical value.
- ***Traffic engineering***. Due to economic considerations, many highways are designed in such a manner that traffic collapse is assumed to take place a limited number (say k) of times during a given period of time. Thus, traffic design is so associated not with the maximum but with the k^{th} largest traffic intensity during that period. Obtaining accurate estimates of the probability distribution of the k^{th} order statistic pertains to the theory of extreme order statistics and allows a reliable design to be made.
- ***Naval Engineering***. Connected quantities in naval engineering are the wave-inducted pitch, roll and heave motions and stresses in ships, and the maxima of these quantities [15].
- ***Corrosion resistance***. Corrosion failure takes place by the progressive size increase and penetration of initially small pits through the thickness of an element, due to the action of chemical agents. It is clear that the corrosion resistance of an element is determined by the largest pits and the largest concentrations of chemical agents and that small and intermediate pits and concentrations do not affect the corrosion strength of the element.
- ***Pollution studies***. The pollution of air, rivers, and coasts has become a common problem for many countries. With the existence of large concentration of people (producing smoke, human waste, etc.) or the appearance of new industries (chemical, nuclear, etc.) The pollutant concentration, expressed as the amount of pollutant per unit volume (of air or water) is forced by government regulations to remain below a given critical level. Thus the regulations are satisfied if, and only if, the largest pollution concentration during the period of interest is less than the critical level.
- ***Meteorology***. Extreme meteorological conditions are known to influence many aspects of human life as in the flourishing of agriculture and animals, the quality of people life, the behavior of some machines, the lifetime of certain materials, etc. In all these cases the engineers, instead of centering interest on the mean values (temperature, rainfall, etc.), are concerned only with occurrence of extreme events (very high or very low temperature,

rainfall, etc.). Accurate prediction of those rare event probabilities becomes the aim of the analysis [12].

1.3. The challenge of extreme meteorological events

There are many different natural phenomena. Some of them are irrelevant in Lithuania because of very rare happenings, or they do not happen at all (e.g. volcanic eruption, tsunami, tornado, earthquake, etc.), while others are relevant because of their significant damage to the environment and humanity.

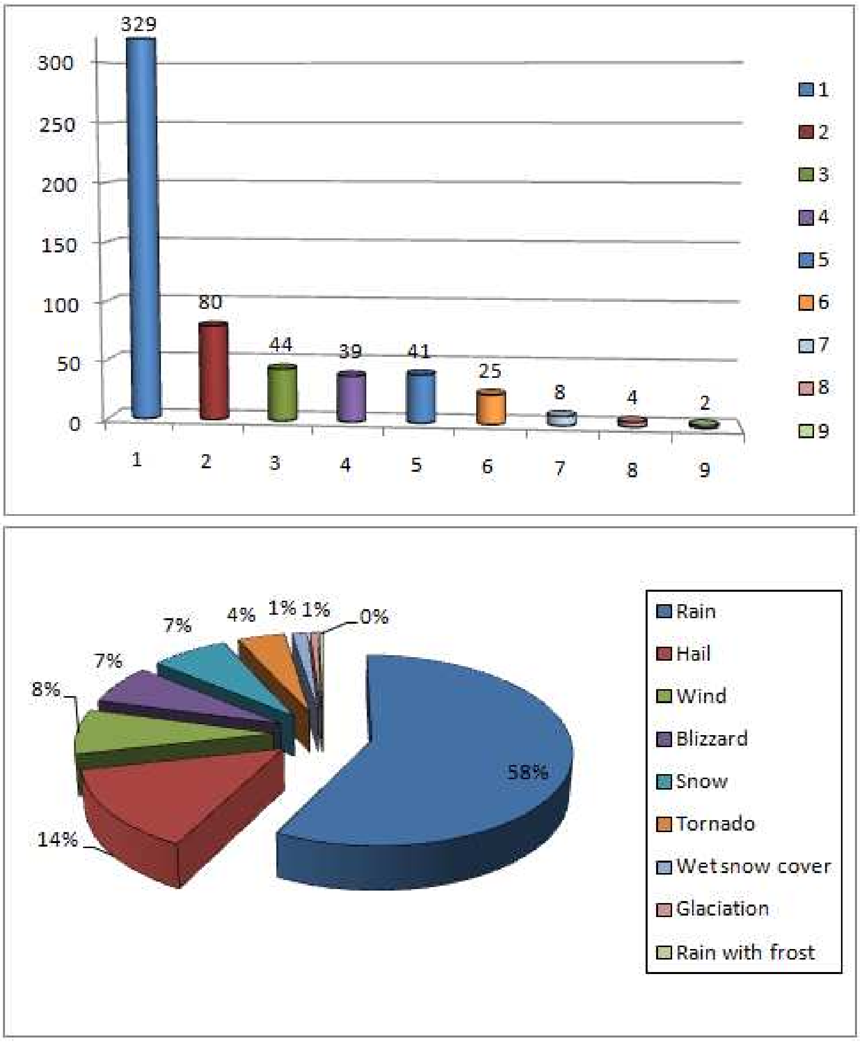


Fig. 3. The number of natural weather phenomena in Lithuania (1961 – 2010) [34]

Weather events in Lithuania are 80 percent naturally occurring emergencies. The rain is the most common meteorological phenomenon that can cause extreme situations (Fig. 3) [22], [23], [24], [25], [26]. Heavy rains averagely occur three to four times a year in Lithuania. Rainfall accessing the criteria for disorganized weather phenomenon causes flash floods and widespread damage, such as: killing people, flooding cultivated fields, destructing the buildings, damaging the equipment. Large hail is 20 mm in diameter pieces of ice and larger precipitations that fall during the warm season in a few or a few dozen minutes in several square miles of land. Such hail can slash slate roofs, break the windows, damage the cars, destroy fruit trees, vegetables, etc. Drought is called a situation where some region of the atmosphere and soil are seeing higher water scarcity than statistically expected for the region. The wind is of relatively horizontal movement resulting from the Earth's surface temperature difference. Heavy rain is one of the most frequently occurring natural phenomena and one of the most dangerous making enormous damage to the humanity in a short period of time. Precipitation is condensed atmospheric water vapor of any form on the Earth's surface. Precipitation is an important part of the hydrological cycle. 505,000 km^3 of water including the 398,000 km^3 falls over the ocean per year.

Extreme precipitation (snow) is a natural phenomenon requiring special attention; we will investigate it further in detail.

Sharp spring flood of 2010 because of ice drifts and heavy snow when a large part of the Nemunas and the Neris coastal towns and villages in Kaunas and Jonava districts were flooded only proves the danger and unpredictability of natural disasters. This is the biggest flood of the last decades in Lithuania.

1.4. Summary of extreme precipitation probabilistic models

The extreme values of generalized algorithms of three main types, such as: I, II, and III (Appendix 3) are used to describe the extreme precipitation.

Gumbel distribution of type I is used to calculate maximum values of extreme weight of snow probabilities over a longer period of time. Gumbel distribution is a single case of Fisher-Tippett distribution, also called the log-Weibull distribution.

The type II of Fréchet distribution is limited from the bottom; shape β parameter is used besides the weight of snow where the scale σ and location μ parameters can be added.

Weibull distribution of type III is most widely used to study weight of snow, since various modifications best meet its particular place and time, also is great for annual rainfall evaluation. Weibull distribution is limited from the top and may have two or three parameters.

All these three types of distribution are easily interchangeable using normal logarithm. A number of other specially adapted to local conditions climatic situations are used in addition to these three basic types of distribution. However, most of them are based on extreme value distributions.

1.5. Uncertainty and sensitivity analysis

Uncertainty and sensitivity analysis is important for studying complex systems as composite laminated structures. Specifically, uncertainty analysis refers to the determination of the uncertainty in response results due to uncertainties in input parameters; sensitivity analysis refers to the evaluation of the contributions of individual uncertainties in input parameters to the uncertainties in response results.

The uncertainty under consideration can be classified as epistemic or aleatory. The epistemic uncertainty is often referred using alternative designations including state of knowledge, subjectivity, and reducibility. The epistemic uncertainty comes from a lack of knowledge of the appropriate value to consider for a quantity that is assumed to have a fixed value used in a particular analysis. Epistemic uncertainty is generally taken to be distinct from aleatory uncertainty under the conceptual and modeling point of view. Aleatory uncertainty arises from inherent randomness in the behavior of the system under study. Designations as variability, stochastic, and irreducible are used for aleatory uncertainty. Several approaches to uncertainty and sensitivity analysis have been developed, including differential analysis, response surface methodology, Monte Carlo analysis, and variance decomposition procedures.

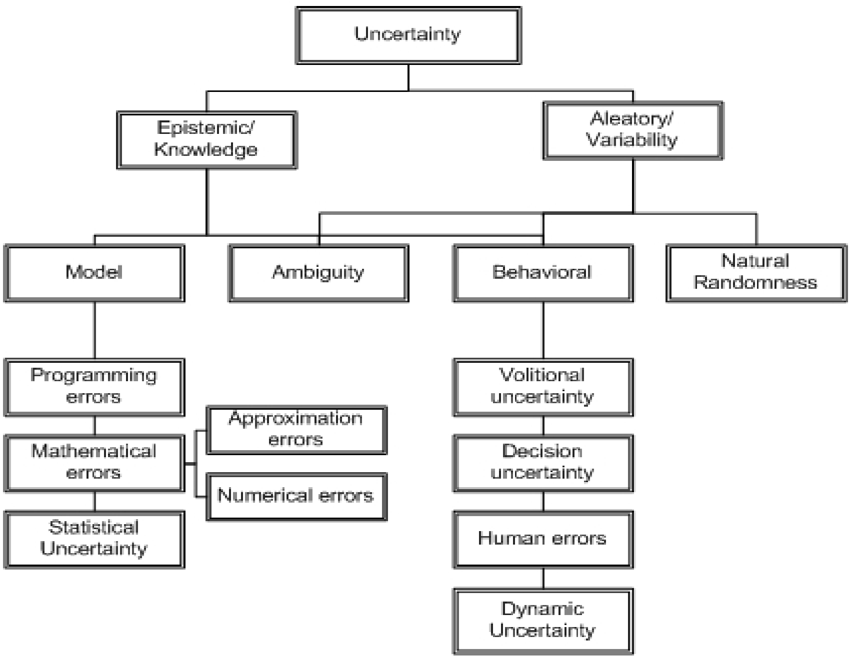


Fig. 4. Uncertainty clasification provided by *CACTUS* [11]

1.6. Probabilistic models of uncertainty in the results

The significance of adequately modeling hydrological extreme events is fully recognized with the increase of both magnitude and frequency of hydrological extreme events, such as drought and flooding. Estimation of extreme snow for various return periods is of prime importance for hydrological design or risk assessment. However, uncertainty involved in extrapolating beyond available data is huge due to knowledge and data limitation.

There are various classifications of uncertainty in the literature. Yen et al. (1986) [16] pointed out that the uncertainties in hydrology and hydraulics include: (1) natural uncertainty associated with inherent randomness of natural process; (2) model uncertainty originated from the inability of model simulation or design procedure to represent precisely the system's true physical behavior; (3) parametre uncertainty resulting from inability to quantify accurately the model inputs and parameters; (4) data uncertainty including measurement errors; and (5) operational uncertainty including human factors that are not accounted in the modeling or design procedure.

Van Asselt (2000) [43] classified the uncertainty based on the modeler's and decision maker's views: model outcome uncertainty and decision uncertainty.

More recently, Walker et al. (2003) [44] have chosen to distinguish three dimensions of uncertainty for uncertainty management in model-based decision support: location, level, and nature of uncertainty. The classification based on the location is mainly identified by logic model formulation, including context uncertainty, model uncertainty, input uncertainty, parametre uncertainty, and model outcome uncertainty. The classification based on the level includes scenario uncertainty, recognized ignorance, and statistical uncertainty. The classification based on the nature includes epistemic uncertainty and variability uncertainty.

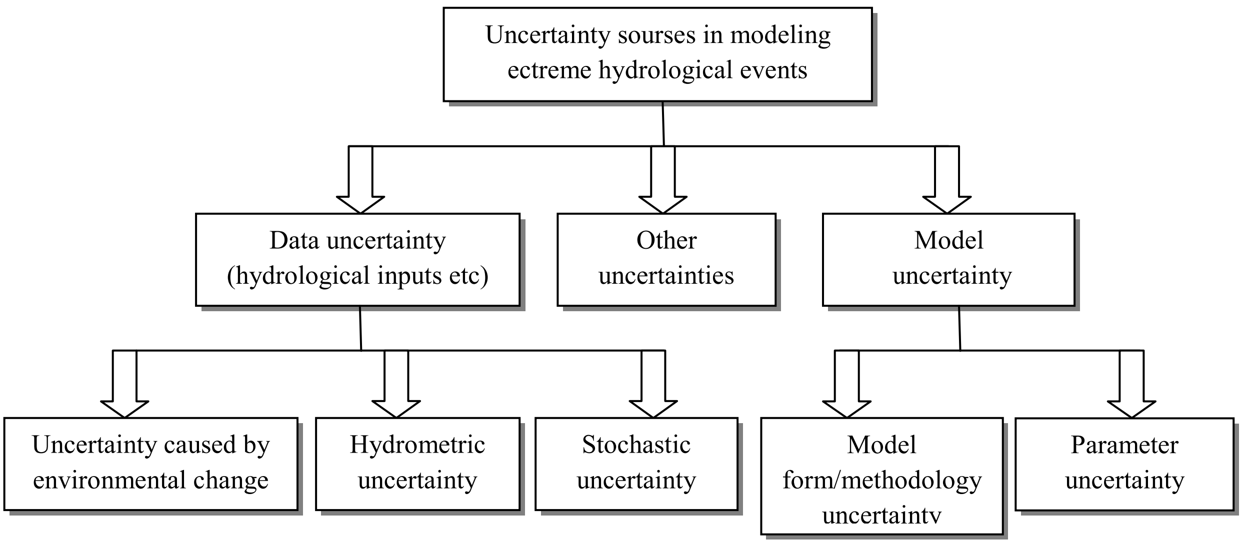


Fig. 5. Classification of uncertainty sources in modeling extreme hydrological events

For the sake of simplicity and clarity, this paper mainly focuses on the modeling issue. Therefore, the uncertainty can be briefly classified into [46]:

1. *Data uncertainty*: mainly including input uncertainty and uncertainty in external factors of the modeled system;

2. *Model uncertainty*: mainly including model form/methodology uncertainty and model parametre uncertainty;

3. *Other uncertainties*: including quantifiable uncertainties, such as: systematic errors and scale-related uncertainty.

Based on this classification, different sources of uncertainties can be identified in modeling extreme hydrological events (Fig. 5, [46]).

1.6.1. Data uncertainty

According to Xu (2009) [46] data uncertainty is closely related to the description of hydrological events and the external variables driving the hydrological system. He distinguishes three main sources of uncertainty:

- the first source is uncertainty caused by environmental changes (climate and land use change);
- the second source of uncertainty results from hydrometric uncertainty in measured hydrological data. For the velocity-area method, the examples of uncertainty may be identified as: instantaneous uncertainty at one point, uncertainty of vertical average velocity calculated using restrict points along the vertical, depth, and velocity sampling error, uncertainty at depth and width measuring and equipment uncertainties, etc.
- the third source of uncertainty is stochastic uncertainty, e.g., uncertainty caused by inherent randomness of natural process or uncertainty resulting from limited sample size during simulation.

The selection of appropriate probability distributions is one central problem in order to quantify data uncertainty. Failure of doing so may result in misleading uncertainty analysis results and therefore lead to poor design decisions. The classical technique of selecting an appropriate input distribution is through goodness-of-fit techniques when sufficient data are available. Modeling data uncertainty can also be implemented through scenario analysis, such as uncertainty caused by climate input, or through simulation by employing Monte Carlo method or others [46].

1.6.2. Model uncertainty

Model uncertainty mainly consists of model form/methodology uncertainty and parameter uncertainty. Three methods: block maximum method, semi-parameter method, and POT method can be used to model extreme hydrological events [46].

The block maximum method is based on the assumption that the data from an independent and identically distributed sample form an exact generalized extreme value distribution asymptotically. In this case Standard statistical methods (maximum likelihood method, method of moments, and L-moment approach [46]) can be used for parameter estimation.

The semi-parameter method is based on a Hill type estimator of the tail index [33]. The POT method is a fully parametric approach based on the generalized Pareto distribution.

The tail estimation following these three different methods gives different values of extreme quantiles. Therefore this type of uncertainty is called model form/methodology uncertainty. In addition to methodology uncertainty, various sources of uncertainty are involved in estimating model parameters.

2. METHODOLOGY

2.1. Probabilistic assessment of extreme weight of snow

2.1.1. Generalized Extreme Value distribution (GEV)

The Generalized Extreme Value (GEV) distribution is a flexible three-parameter model that combines Gumbel, Fréchet, and Weibull *maximum* extreme value distributions, also known as type I, II, and III extreme value distributions, with cumulative probability distribution function given by [10]:

$$F(x) = \exp\left\{-\left[1+\frac{\beta(x-\mu)}{\sigma}\right]^{-\frac{1}{\beta}}\right\}, \tag{2.1}$$

$$1+\frac{\beta(x-\mu)}{\sigma} > 0, \tag{2.2}$$

where μ, σ, and β are the location, scale, and shape parameters respectively. The scale must be positive $(\sigma > 0)$, the shape and location can take on any real value.

The classical statistical analysis is based on the assumption that the cumulative probability distribution F of the observed data X does not vary over time and the results are obtained using maximum likelihood estimation. The approximation may remain valid under some regularity conditions, thus extending the framework of the GEV to modeling of annual maxima M_j of year j by a generalized extreme value distribution depending on the parameters varying from year to year. If F is considered to be non-constant for each year, but constant throughout each individual year. In that case the independent variables with a non-stationary GEV distribution are given by [10]:

$$F(x) = \begin{cases} \exp\left(-1+\beta\left(\frac{x-\mu(j)}{\sigma(j)}\right)^{\frac{-1}{\beta}}\right) & if\, \beta \neq 0 \\ \exp\left(-\exp\left(-\frac{x-\mu(j)}{\sigma(j)}\right)\right) & if\, \beta = 0 \end{cases}. \tag{2.3}$$

The shape parameter β is time invariant. The shape parametre is considered to be the main one describing the behavior of the probability distribution at extreme levels.

2.1.2 Gumbel distribution

The Gumbel probability distribution function [6] is

$$F(x) = P(X < x) = e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}, -\infty < x < \infty; \sigma > 0, \tag{2.4}$$

where μ and σ are the constants known as the location and scale parameters, respectively. In this case only two parameters (the equality of mean and variance) exist. The moment equation becomes [6]:

$$\bar{x} = \mu + 0.5772\sigma, \tag{2.5}$$

$$s_x^2 = \frac{\pi^2 \sigma^2}{6}, \tag{2.6}$$

where $\bar{x}$ and s_x^2 are the sample mean and quasi-variance. From (2.5) and (2.6) the following moment estimates [6]:

$$\sigma = \frac{s_x \sqrt{6}}{\pi}, \tag{2.7}$$

$$\mu = \bar{x} - 0.5772\sigma \tag{2.8}$$

are obtained.

Because probability analysis is related to extreme values, the survival function $G(x)$:

$$G(x) = P(X > x) = 1 - e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}, -\infty < x < \infty, \sigma > 0 \tag{2.9}$$

is used.

It is reasonable to adjust functional expression of Gumbel extreme precipitation distribution, because follow-up analysis will be linked to estimation of mean and standard deviation. The main parameters μ and σ correspondingly are replaced by the estimate of average $\bar{x}$ and standard deviation s_x. Hence adjusted Gumbel model is

$$\mathrm{G}(x, \bar{x}, s_x) = 1 - \mathrm{e}^{-\mathrm{e}^{-\pi\left(x - \left(\bar{x} - 0.5772\frac{s_x\sqrt{6}}{\pi}\right)\right)/s_x\sqrt{6}}}, \tag{2.10}$$

where x is the level of weight of snow, $\bar{x} = 78.778$ and $s_x = 64.436$.

2.1.3 Weibull distribution

The three-parameter Weibull distribution is given by the distribution function [29]:

$$F(x) = P(X < x) = 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^{\beta}}, x \geq \mu, \tag{2.11}$$

where μ, σ, and β are the constants known as the location, scale, and shape parameters, respectively and such that $\sigma > 0$ and $\beta > 0$. The main parameters of Weibull distribution are evaluated using threshold parameter. Transforming Weibull cumulative distribution function (CDF) to [5]:

$$\log(x-\mu)=\frac{1}{\beta}\log(-\log(1-p))+log\sigma \tag{2.12}$$

gives a linear relationship between $\log(x-\mu)$ and $\log(-\log(1-p))$ where μ is a threshold parameter and p is a cumulative probability. Recording threshold parameter and using the least squares the shape and scale parameters could be found. The parameter of determination R^2 value of a linear regression on the transformed variables $\log(x-c)$ and $\log(-\log(1-p))$ is needed to be maximized in order to evaluate the location parameter μ.

2.1.4 Fréchet distribution

The Fréchet CDF is given by [6]:

$$F(x)=P(X<x)=e^{-\left(\frac{\sigma}{x-\mu}\right)^{\beta}},x\geq\mu, \tag{2.13}$$

where μ, σ, and β are the constants known as the location, scale, and shape parameters, respectively and such that $\sigma>0$ and $\beta>0$. Fréchet parameter estimates are obtained with the help of maximum likelihood estimates.

2.1.5 Return period

According to Shiau (2003) [39] an extreme event occurs if the random variable $X_i, i=1,\ldots,n$ is greater or equal to some magnitude x_T. The recurrence interval T is defined as time period between occurrences for the event $X_i\geq x_T$. The return period for the event $X_i\geq x_T$ is the expected value of T, denoted by $E(T)$ in this study. The return period for an event of a given magnitude is thus defined as the average reccurence interval between events equaling or exceeding a specific magnitude [39].

The return period for the event $X_i\geq x_T$ can be related to the probability of occurrence for such events in the following way. It is assumed that the probability of occurrence for the event $X_i\geq x_T$ in any year is $P(X_i\geq x_T)$. X_i is serially independent in the values, the probability that time interval T between exceedance of a precipitation magnitude x_T equals n is given by [21]:

$$P(T=n)=P(X_1<x_T)P(X_2<x_T)\ldots P(X_{n-1}<x_T)P(X_n\geq x_T)=P(X_1<x_T)^{n-1}P(X_n\geq x_T). \tag{2.14}$$

The expected value for T is then given by [21]:

$$E(T)=\sum_{n=1}^{\infty}nP(X_{n-1}<x_T)^{n-1}P(P(X_n\geq x_T))=\frac{1}{P(X_i\geq x_T)}=\frac{1}{1-P(X_i<x_T)}. \tag{2.15}$$

The above equation is the return period for an extreme event described by a single random variable and based on the annual maximum series. Obviously, the return period depends on the distribution of the selected random variables, i. e., the longer the return period $E(T)$, the less the frequency $P\left(X_i \geq x_T\right)$, and the larger the magnitude of the random variable x_T.

2.2 Model selection and validation

Model selection is a difficult task and the choice of a suitable model is made on the basis of whatever knowledge is available and with the use of well-considered judgment. It is important that the selected model would be enough flexible enough to model the data adequately, taking into account the compromise between ease of analysis and the complexity of the model.

There are two ways for selection of the model to the analyzed data: *graphical models* and *goodness-of-fit* test [29].

Graphical methods of model selection involve preparation of probability plots or other graphs and check visually in order to determine whether or not plots appear to show certain characteristics of a particular model. Nelson (1982) and O'Connor (1997) discuss this issue more detaily [29].

Goodness-of-fit tests are statistical procedures for testing hypothesised models. According to Nelson a poor fit (graphical or analytical) may occur for two reasons: (1) the model is incorrect, or (2) the model is correct, but the parameter values specified or estimated may differ from the true values by too great amount. In general, validation requires testing, additional data and other information, and careful evaluation of the results.

2.2.1 Graphical models

Murphy, Xie, and Jiang (2004) [29] have introduced graphical models as a tool for visualizing fitness of data to one or more of the models by plotting data. Visual fits provided by graphical methods are quite useful. The benefits of the graphical approach are the visual insights and (usually) ease of application.

They provided several main problems with the graphical approach that are related with a) subjective (to some extent) and b) no well-developed statistical theory for determining the small sample or asymptotic properties of the procedure. According to them the first problem is not serious; it means that different analysts are likely to arrive at somewhat different results if the plotting is done by hand or using different algorithms, but they usually will come to the same conclusions. The second problem is more serious; it means that standard errors and distributions of estimators are not known, even asymptotically, and that test statistics cannot be obtained. They concluded saying that proper analysis of a set of data cannot be based exclusively on graphical procedures, and one needs to use statistical and analytical methods as well.

There are various kinds of plot used for comparison of different extreme value distributions [29]:

1. ***Probability plot***. The EDF plot (for complete data) involves plotting $\hat{F}(X_{(i)})$ versus $X_{(i)}$ ($X_{(i)}, i = 1, \ldots, n$ is the data set) with $\hat{F}(X_{(i)})$ computed as indicated earlier. Without loss of generality, assume that T is a standardized random variable and $F(X)$ a standardized distribution. A probability plot is a plot of $z_i = F^{-1}\left(\hat{F}(X_{(i)})\right)$ versus $X_{(i)}$. If F is true CDF, the probability plot is approximately a straight line.
2. ***Weibull plot.*** The plotting depends on the type of data. It involves plotting of y_i versus x_i given by:

$$y_i = \ln\{-\ln[1 - \hat{F}(X_{(i)})]\} \text{ and } x_i = \ln\left(X_{(i)}\right). \quad (2.16)$$

   If the plotted data is roughly along a straight line, then one can model the data by the standard Weibull model. If the plot is not a straight line, then one or more models depending on the shape derived from the standard Weibull model might adequately model the given data set.
3. ***P–P Plot.*** A P–P is a plot of percentages of one distribution versus that of a second distribution. Wilk and Gnanadesikan (1968) [14] stated that they are useful for detecting discrepancies in the middle of the distribution (about the median) while the P–P plots are limited in their usefulness.
4. ***Q–Q Plot.*** A Q–Q plot is a plot of the quantiles (or percentiles) of one distribution versus the quantiles of a second distribution. If one of these distributions is a hypothesized theoretical distribution, then a Q–Q plot is just a probability plot discussed earlier (Wilk and Gnanadesikan (1968) [14].

2.2.2 P–P and Q–Q plots

Rinne (2009) [36] specified that both types of plots are suitable to compare as:

- two theoretical CDFs,
- two empirical CDFs

or

- an empirical CDF to a theoretical CDF.

Let $F_X(x)$ and $F_Y(y)$ be the CDFs of the variates X and Y, respectively, given in Fig. 6 [36]. Two types of graphs (the Q–Q plot or quantile plot and the P–P plot or percent plot) were deduced from this display.

For each P value on the ordinate axis displaying the CDF, there are most two values on the abscissa axis displaying the realizations of the variates, called quantiles:

$$x_P = Q_X(P) \text{ and } y_P = Q_Y(P).$$

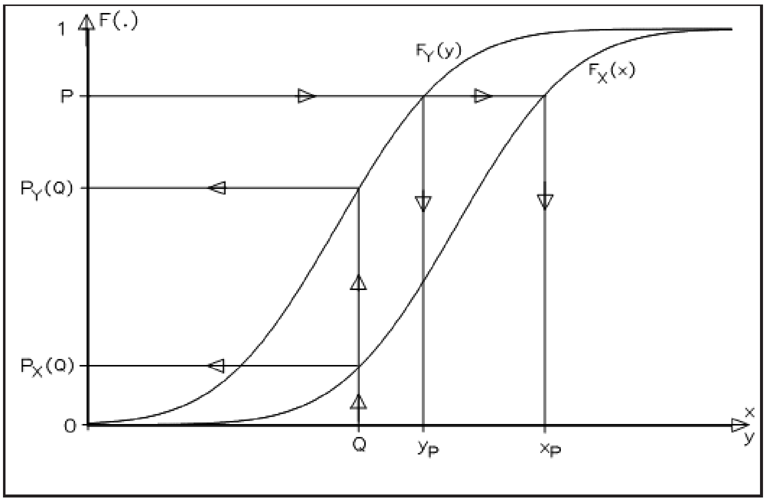


Fig. 6. Explanation of the Q–Q and the P–P plots

A ***Q–Q plot*** is a display where y_P is plotted against x_P with P varying $0 < P < 1$. Conversely for each value Q on the abscissa axis there are mostly two values on the ordinate axis:

$$P_X(Q) := F_X(Q) \text{ and } P_Y(Q) := F_Y(Q).$$

A ***P–P plot*** is a display where $F_Y(Q)$ is plotted against $F_X(Q)$ for Q varying, $Q \in \mathrm{R}$. There are several modifications of these two basic displays. There is a special case where Q–Q plot and P–P plot are identical: X and Y both are uniformly distributed. This will be the case when two variates are probability integral transforms.

The P–P plot is less important than Q–Q plot. If X and Y are identically distributed, their CDFs will coincide in Fig. 6 and the resulting P–P plot will be a 45^{o}–line running from $(P_X(Q), P_Y(Q)) = (0,0)$ to $(P_X(Q), P_Y(Q)) = (1,1)$. On the contrary to the Q–Q plot, the P–P plot will not be linear if one of the two variates is a linear transform of the other one.

2.2.3 P–P plot for the Gumbel model

The P–P plot in conjunction with Gumbel distribution function $F_{\mu,\sigma}$ is given by Reiss and Thomas (2007) [35] as:

$$\left(q_i, F\left(\left(x_{i:n} - \mu_n\right)/\sigma_n\right)\right), i = 1, \ldots, n, \tag{2.17}$$

where μ_n and σ_n are estimates of the location and scale parameters.

Because

$$F\left(\left(x_{i:n} - \mu_n\right)/\sigma_n\right) = F_{\mu_n, \sigma_n}\left(\hat{F}_n^{-1}\left(q_i\right)\right), \tag{2.18}$$

a strong deviation of the P–P plot from the main diagonal in the unit square indicates that the given model is incorrect (or the estimates of the location and scale parameters are inaccurate). The values of P–P plot will be close to one (or zero) and, thus close to the diagonal in the upper (or lower) tail, even if the choice of the model is wrong.

2.2.4 Q–Q plot for the Gumbel model

Assume that the data $x_1, \ldots, x_n$ are governed by a distribution function [35]:

$$F_{\mu,\sigma}(x) = F\left(\left(x - \mu\right)/\sigma\right) \tag{2.19}$$

with location and scale parameters μ and $\sigma > 0$. Thus $F = F_{0,1}$ is a standard version. Values $\hat{F}_n^{-1}\left(q\right)$ of the sample distribution function will be plotted against $F^{-1}(q)$. More precisely, one plots the points [35]:

$$\left(F^{-1}\left(q_i\right), \hat{F}_n^{-1}\left(q_i\right)\right), i = 1, \ldots, n, \tag{2.20}$$

where $q_i = \dfrac{i}{n+1}$. The location and scale parameters must not be selected in advance when a Q–Q plot is applied to the data. Because $\hat{F}_n^{-1}\left(q_i\right) = x_{i:n}$, the relationship [35] is

$$x_{i:n} = \hat{F}_n^{-1}\left(\frac{i}{n+1}\right) \approx F^{-1}\left(\frac{i}{n+1}\right). \tag{2.21}$$

Between the sample guantile function F^{-1} and the underlying quantile function yields [35]:

$$\hat{F}_n^{-1}\left(q_i\right) \approx F_{\mu,\sigma}^{-1}\left(q_i\right) = \mu + \sigma F^{-1}\left(q_i\right), \tag{2.22}$$

and, hence, the Q–Q plot of points:

$$\left(F^{-1}\left(q_i\right), x_{i:n}\right), i = 1, \ldots, n \tag{2.23}$$

is closed to the graph $\left(x, \mu + \sigma x\right)$. The Q–Q plot can be visualized by a scatterplot, whereby a linear interpoliation may be added. Apparently, the intercept and the slope of the Q–Q plot provide estimations of μ and σ. Another frequently taken choice of q_i is $\dfrac{i-0.5}{n}$ [35].

The selected location/scale parameter family (Gumbel model) is untenable if the deviation of the Q–Q plot from a straight line is too strong.

2.2.5 Probability Difference Graph

The probability difference graph is a plot of the difference between the empirical CDF [8]:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x_i \le x) \tag{2.24}$$

and the theoretical CDF:

$$Diff(x) = F_n(x) - F(x). \tag{2.25}$$

This graph is used to determine how well the theoretical distribution fits to the observed data and to compare the goodness of fit of several fitted distribution. It is displayed as a continuous curve or a scatterplot for continuous distributions and a collection of vertical lines for discrete distributions.

2.2.6 *Goodness-of-fit* tests

The *goodness-of-fit* (GOF) tests measure the compatability of a random sample with a theoretical probability distributions function. In other words, these tests show how well the selected distribution fits to the particular data.

A general test of fit is a test of a null hypothesis:

H_0 : the given data comes from a CDF $F(X;\mu,\sigma,\beta)$;

with the form of the CDF specified. One needs to consider the following cases [29]:

Case 1: The parameters $\mu,\sigma,$ and β are fully specified (or known).

Case 2: The parameters $\mu,\sigma,$ and β are not fully specified (or known) and the unspecified parameters need to be estimated from the given data.

The theory is well developed for case 1, the data set is complete and give appropriate references for the test for incomplete data.

The basic idea in testing H_0 is to look at the data obtained and to evaluate the likelihood of occurrence of this sample given that H_0 is true. H_0 is rejected if the conclusion is that this is highly unlikely sample under H_0. It follows that two types of errors can be made: a) rejecting H_0 when it is true (type I error), and b) failing to reject H_0 when it is not true (type II error). The level of significance (denoted by α) is the maximum probability of making a type I error and the power of the test is the probability of not making the type II error. The test statistic is a summary of the given data, and the test is usually based on such statistics. The test involves comparing the computed values of the test statistic with some critical values. The hypothesis is rejected if the statistics exceeds the critical value [29].

A set of extreme weight of snow data is completed and the parameters μ, σ, and β are fully specified therefore these laters allow us to be familiar only with the case 2. There are three main tests in order to evaluate the best fit of extreme weight of snow to any extreme value distributions [29], [41]:

- Chi-Square Test;
- Kolmogorov-Smirnov Test;
- Anderson-Darling Test.

2.3 Methodology of uncertainty and sensitivity analysis

2.3.1 Simulation of events

It is necessary to have models reflecting key features of real elements in order to simulate natural processes. The results would reflect the processes taking place in real space; mathematical model would be as simple as possible, event modeling is divided into physical and statistical, physical modeling is based on the solution of mathematical equations describing analyzed processes. Long-term events characteristic measurement data and statistical relationship between prognosis and actual data are performed during statistical analysis. It is convenient to describe mathematical models as a function:

$$y = F(x_1, x_2, \ldots, x_N). \tag{2.26}$$

Mathematical model description (2.33) is slightly simplified, because realistic models usually have not only one but a lot of results, describing different characteristics of the analyzed process or phenomenon, i.e. y is a vector, but not a scalar. It is also necessary to analyze the results depending on time t, i.e. $y=y(t)$, and the function F also depends on t. However, a simplified model description (2.33) does not change the principles of uncertainty and sensitivity analysis.

The model function $F(\cdot)$ linking model parameters to the result y in common case is complex and often a non-linear function with a rarely known analytical expression. Function $F(\cdot)$ can be used as calculations describing function in complex models using specialized software packages. Important facilitation is that the uncertainty and sensitivity analysis does not need to know the analytic $F(\cdot)$ expression.

2.3.2 Methodology of uncertainty analysis

Uncertainty analysis is an attempt to quantify the degree of confidence that an analyst has in the existing data and models, based on whatever information is currently available. Estimates of risk and reliability models are becoming central to decisions about many engineering systems. However, we do not have enought data in many important cases. Rare events are prime examples. It is important to estimate the uncertainty as quantitative measure, e.g. probability when data is scarce or physical understanding is limited. It is possible to treat model parameters as random variables or as the parameters of the probability distributions of the random variables in order to estimate the uncertainty as probability. The uncertainty analysis models already are very highly applied in nuclear engineering field.

In present research real time snow forecast model is the most important and having the biggest uncertainty. This section presents uncertainty and sensitivity analysis methodology, performed for snow forecast model results and parameters.

The scheme of the link between model parameters uncertainty and modelling results uncertainty is presented in Fig. 7 [19].

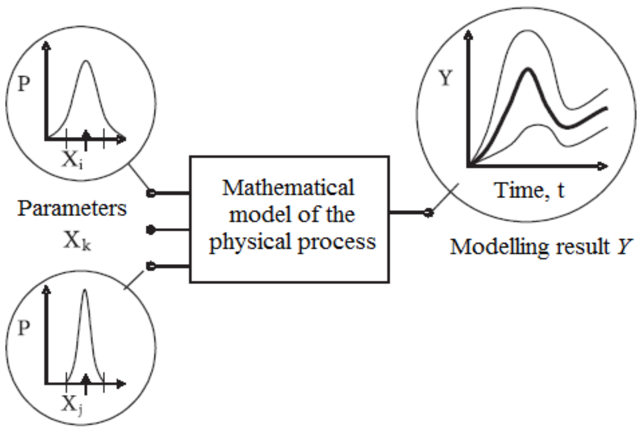


Fig. 7. Link between uncertainty of model parameters and uncertainty of modeling results

The uncertainty analysis of the results was performed using the popular software SUSA (Software System for Uncertainty and Sensitivity Analysis) [18]. In order to use its standard procedures the following steps must be performed:

1) The determination of potential uncertain model parameters, which are important for modelling results;
2) The determination of uncertainty limits for each parameter. The distribution shape, minimal and maximal parameter values are usually determined from the literature, experience, engineering judgement or other reasonable sources;
3) The determination of links between uncertain parameters. It is important to note that sample generation and post-processing is considerably more complicated at presence of correlated parameters. Tipically, analysts prefer to selected uncorrelated parameters or their combinations;

4) The generation of random samples of considered parameters and simulations. The number of generated parameters values is estimated using S.S. Wilks formula. Wilks method enables to determine the minimal number n of model simulations that the obtained minimal and maximal values would agree with considered level and confidence interval limits, i.e. for tolerance interval (α, β) estimation, n must satisfy the following inequality [45]:

$$1 - \alpha^n - n(1-\alpha)\alpha^{n-1} \geq \beta, \tag{2.27}$$

where α – limit parameter; β – confidence level.

Table 1. Minimal numbers of model simulations

	(α, β) – tolerance interval				**(α, β) – tolerance limit**		
β/α	**0,90**	**0,95**	**0,99**	**β/α**	**0,90**	**0,95**	**0,99**
0,90	38	77	388	**0,90**	22	45	230
0,95	46	93	473	**0,95**	29	59	299
0,99	64	130	662	**0,99**	44	130	459

Using this formula the two-sided tolerance interval of (95%,95%) gives n=93 (Table 1).

5) Evaluation of parameter uncertainties using simulation results. The uncertainty of the final result is determined by a probabilistic distribution, depending on imput parameters distribution. The two-sided tolerance interval is commonly used (95%,95%), and interpreted as follows: at least $\alpha \cdot 100\%$ of all model results will be contained within the tolerance interval with a probability of at least 0.95.

Tipically it is very efficient for large models to use the same uncertainty analysis sample to perform also sensitivity analysis. The sensitivity analysis enables to determine parameters which are the major contributors to model result uncertainty. The most commonly used sensitivity indexes are correlation coefficients, standardised regression coefficients and others.

Pearson's and Spearman's rank correlation coefficients were used in present analysis. The parameter x_i correlation coefficient determines the linear dependency between parameter and model result. The correlation coefficient is estimated using data ranks determining the influence level of parameter on the modelling results when the dependency between parameter and result is not linear.

2.3.3 Methodology of sensitivity analysis

The purpose of sensitivity analysis is twofold; the first is to set the system simulation sensitivity results from initial parameters and the second is the analysis of key modeling assumptions for final results. Categorization of the parameters according to their relative contribution to the whole uncertainty of the result and quantitative assessment of the contribution to each parameter is the most commonly used quantitative sensitivity analysis of uncertainty.

One of the main goals of sensitivity analysis is to determine a model with more accurate assessment for significant model result reduce. The uncertainties and further adjustment of parameters are not meaningful because of their low impact on the result. Such evaluation is important in order to:

- evaluate the possibility of model application;
- determine the most significant parameters in order to obtain more accurate model results;
- understand the basic functioning dependencies of the modeled system.

The choice of sensitivity analysis method depends on:

- choice of sensitivity assessment;
- desired sensitivity assessment accuracy;
- model research costs and others.

In general, sensitivity analysis and its application highly depend on choice of the sensitivity assessment. Most of the sensitivity assessments are applied with mathematical, simulation, and other models that can be expressed in the following form:

$$F\left(u\left(n\right),k\left(m\right)\right)=0,$$

where k is a set of parameters of m size and u is n size set of results.

According to the choice of sensitivity metrics and possible model parameter variations, the sensitivity analysis methods can be classified into the following categories:

- ***General model parameters or structure variation***. The results are calculated besides different parameter combinations or the direct model structure (including minuteness) amendments. Sensitivity estimates (that are usually approximate) in this case are determined by random and freely chosen values of parameters, including the nominal and extreme parameter values. The most important is the sensitivity estimates based on extreme values of the system model parameters during the system safety analysis.
- ***Defined field*** sensitivity analysis. Functioning of the system; in this case model parameters are analyzed varying in pre-defined boundaries that are often chosen depending on the analysis of uncertainty in the parameters set by the uncertainty limits.
- ***Local*** sensitivity analysis. Sensitivity of model results is studied slightly varying about the single parameter set. Local sensitivity analysis results are often characterized using gradients and partial derivatives besides the selected parameters profile.

The sample and dispersion fragmentation methods are used in sensitivity analysis. Standardized linear regression is one of the most popular sampling methods. The mathematical model (2.27) is expressed by multiple linear parametric function:

$$y = F\left(x_1, x_2, \ldots, x_N\right) = \alpha + b_1 x_1 + \ldots + b_N x_N. \tag{2.28}$$

where b_i are calculated by the least square method, but they can not be sensitive indices as parameter measurement scales need to be normalized. Parameters measuring units are normed by standardizing each unit and model result:

$$\hat{x}_{i,k} = \frac{x_{i,k} - Ex_i}{\sigma x_i}, i = 1,2,\ldots,N;\ k = 1,2,\ldots,M; \tag{2.29}$$

$$\hat{y}_k = \frac{y_k - Ey}{\sigma y}, \tag{2.30}$$

where:

$Ex_i - x_i$ is the average parameter;

Ey – the mean of model result;

σx – parameter standard deviation;

σy – standard deviation of model result;

M – random sample of size parameters;

N – number of parameters.

The regression coefficients β_i for standardized values are called standardized regression coefficients, and parameters sensitivity indices:

$$\hat{y} = \alpha + b_1 \hat{x}_1 + \ldots + b_N \hat{x}_N \tag{2.31}$$

are used there.

Standardized regression coefficient results are based on model linearity hypothesis. It is important to calculate the linear model coefficients in order to confirm this hypothesis

$$R^2 = \frac{\sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{p} (y_i - \overline{y})^2} \tag{2.32}$$

where $\hat{y}_i$ is the y_i estimate from the regression model.

R^2 shows how exactly a linear regression model reflects the actual Y receipt. The model accuracy is higher, the R^2 is closer to 1. In the other words, such sensitivity means 'correlation coefficient' or 'standardized regression coefficient' explaining the relatively small sample values Y part, if, for example $R^2 < 0.5$. In practice it is often required that the linear model's coefficient of determination is not less than 0.6, i.e. parameter uncertainties must explain not less than 60%

uncertainty of model result. Correlation coefficients can not be parameter sensitivity indexes at low R^2.

2.4 Uncertainty analysis software

SUSA is a Software System for Uncertainty and Sensitivity Analysis and it was developed by GRS [18]. GRS method together with the software package SUSA typically presents a quantitative method used for uncertainty and sensitivity analysis in verified statistical methods.

A new simulation project is being created in order to make model uncertainty and sensitivity analysis using SUSA. The user makes the list of the uncertain parameters used in modeling project and provides all the information that is used to describe the existing knowledge of each parameter. There is a possibility to insert or remove undetermined parameters at any time.

The user can select the type of information to be described for each parameter:

- to copy the available knowledge of the current parameter settings assigned to the other parameter and, where appropriate, to make changes;
- to indicate that would like to define exactly the dependency information (conditional distribution), as the current parameter distribution depends on the availability of knowledge about the other parameter description;
- to enter the value area and other distribution characteristics (such as distribution type and parameters or quintiles) for current parameter.

The user can create the parameter value sample when the obtained knowledge is evaluated quantitatively. The following two types of sample selection procedures are introduced:

- simple random sampling;
- Latin hypercube sampling (can be chosen between conditional median and a random selection value from each subinterval choice).

SUSA gives a number of model output uncertainty study possibilities: the user can specify the calculation of empirical quartiles and permissible strips; Lilliefors and Kolmogorov-Smirnov tests are performed to choose model output empirical distribution for parametric distributions. There is a possibility to perform graphical output. Chart shows the empirical distribution function of the model output together with the permissible limits (if needed) or chosen parametric distributions (if needed). It is a possibility to compose approximate simplified linear model, related with uncertain parameters.

SUSA presents four correlation coefficient groups related with the following sensitivity of the indicators:

- Pearson's;

- Blomquist's;
- Kendall's;
- Spearman's.

Simple and partial correlation as well as standardized partial regression and determination R^2 coefficients are calculated in each group. In addition, user can create sensitivity characteristics from the correlation relations, 2x2 feature dependence tables or rank regression, and sensitivity indicator charts.

SUSA at first generates graphs from the values, stored in the design file and / or in model output file. SUSA firstly derives the output file, identified by submitting the different pairs of correlation coefficient values and then displays the required resolution graphics in the title by a correlation coefficient values. In addition, the relationships between ranked parameters of the sample values (model results) and ranked sample model results field values (parameters).

3 INVESTIGATORY PART

3.1 Probability assessment for several extreme value distributions

Impact of snow weight on any structures depends on snow thickness and density. Weight of snow is calculated by multiplying the thickness and density of it and can be expressed in terms of snow water equivalent.

Estimates of extreme weight of snow used in further calculations are derived from annals of Lithuanian Hydrometeorological Service and Dūkštas MS. The snowfall data consist of annual maximum weight of snow values in Dūkštas region during the period of 1992-2009. This region of Lithuania is chosen because of importance of Ignalina NPP. The probability of maximum weight of snow and the return period are the main parameters in probability assessment. The time series of these data for the period of 18 years are shown in Fig. 8 (Appendix 2).

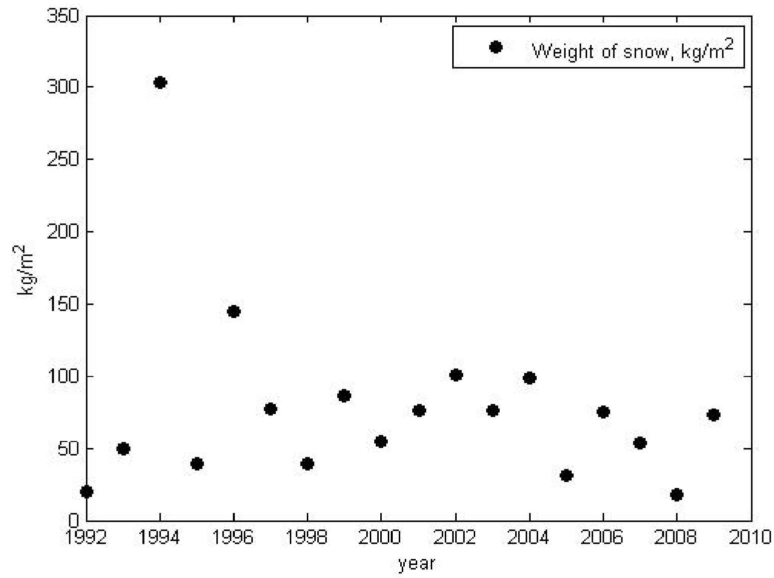


Fig. 8. Scatter plot of the annual maximum weight of snow data in Dūkštas MS

Further the dependence of the exceedance probability of extremes on the various extreme value distributions will be analysed.

Table 2. Location, scale, and shape parameters of extreme value distributions

	Gumbel	Weibull	Fréchet	GEV
μ	49.78	13.64	-66.10	50.74
σ	50.24	68.34	116.84	30.40
β	-	1.26	3.84	0.26

Parameters of Gumbel, Weibull, Fréchet, and GEV distributions which were estimated by using moment method, threshold estimates, and maximum likelihood method are presented in Table 2.

The probability that extreme weight of snow exceeds more than 50-300 kg/m^2 is calculated by using frequencies of extreme weight of snow together with probabilistic model (Table 3).

Table 3. Probabilities of exceedance under various models

The annual extreme weight of snow, kg/m^2	Gumbel	Weibull	Fréchet	GEV
50	$6.306\cdot10^{-1}$	$6.360\cdot10^{-1}$	$6.421\cdot10^{-1}$	$6.412\cdot10^{-1}$
70	$4.877\cdot10^{-1}$	$4.562\cdot10^{-1}$	$4.272\cdot10^{-1}$	$4.267\cdot10^{-1}$
100	$3.079\cdot10^{-1}$	$2.613\cdot10^{-1}$	$2.283\cdot10^{-1}$	$2.280\cdot10^{-1}$
150	$1.272\cdot10^{-1}$	$9.23\cdot10^{-2}$	$9.05\cdot10^{-2}$	$8.98\cdot10^{-2}$
200	$4.905\cdot10^{-2}$	$2.94\cdot10^{-2}$	$4.16\cdot10^{-2}$	$4.14\cdot10^{-2}$
250	$1.842\cdot10^{-2}$	$8.6\cdot10^{-3}$	$2.05\cdot10^{-2}$	$2.16\cdot10^{-2}$
300	$6.849\cdot10^{-3}$	$2.3\cdot10^{-3}$	$1.25\cdot10^{-2}$	$1.23\cdot10^{-2}$

Calculated maximum likelihood estimates of probabilities of annual extreme weight of snow are shown in Fig. 9. Estimated values of Gumbel and Weibull distributions are similar as well as values of Fréchet and GEV distributions respectively (Table 3 and Fig. 9).

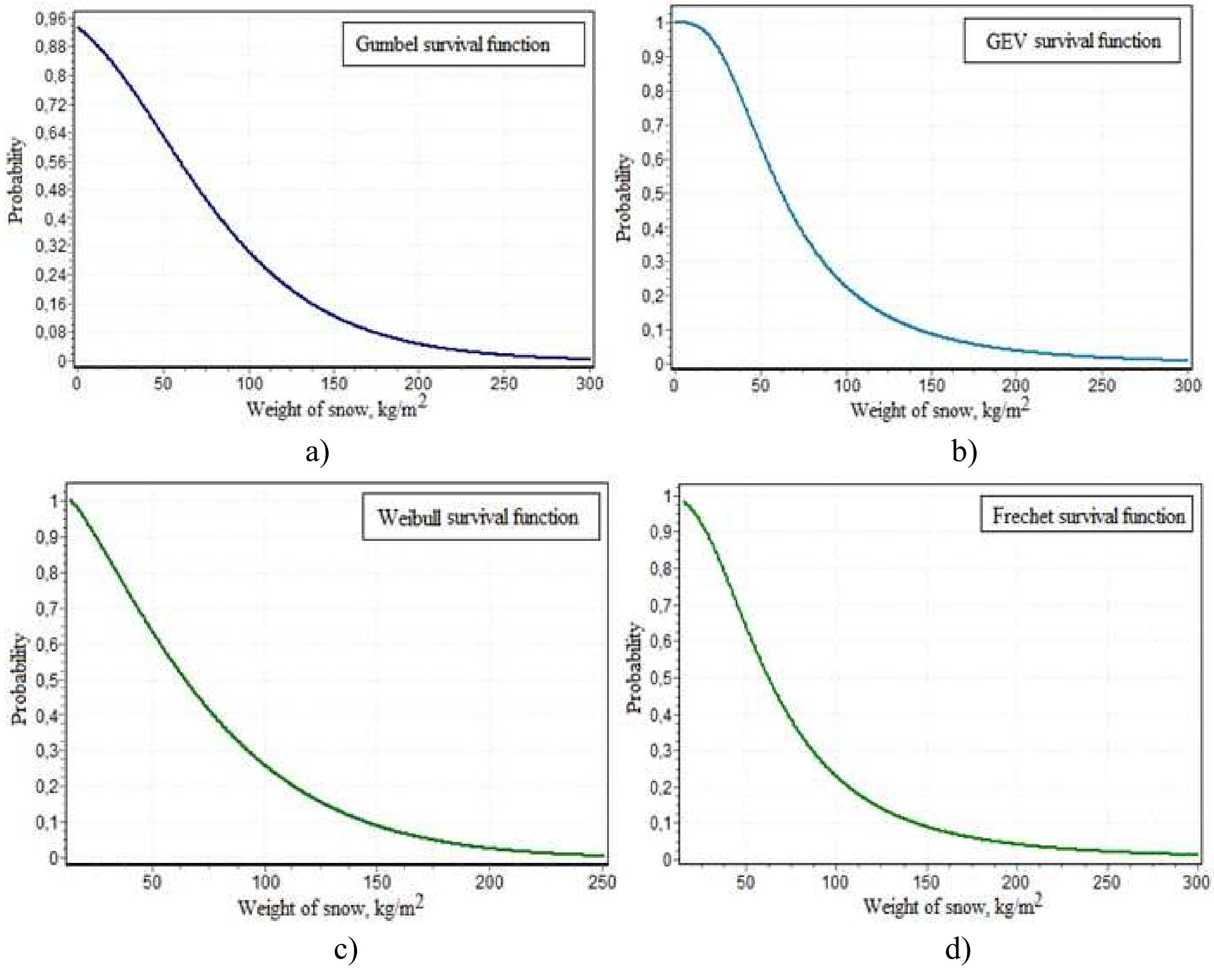


Fig. 9. Occurrence probability per year: a) Gumbel, b) GEV c) Weibull, and d) Fréchet

Performed probabilistic assessment might be used further to analyse the uncertainty and sensitivity of estimates of extreme snow weights; this allows evaluation of uncertainties present in extreme value modelling above. Also calculated uncertainty limits might be used for the decision of which distribution has the best fit for statistical data.

3.2 Return period of extreme weight of snow

It is necessary to have the relationship between extreme value distribution and return period in order to estimate extreme values x with particular return period. From Gumbel (G) survival function $G(x)$ (2.9) and from definitions of return period $E(T)$ we can write:

$$1 - e^{-e^{-y}} = \frac{1}{E(T)}. \quad (3.1)$$

Hence

$$y = -\ln(\ln\left(E(T)\right) - \ln(E(T) - 1)), \text{ where } \quad y = \frac{x - \mu}{\sigma}, \quad (3.2)$$

$$x_G = \mu + \sigma\left(-\ln\left(\ln\left(E(T)\right) - \ln\left(E(T) - 1\right)\right)\right). \quad (3.3)$$

According to other relationship between Weibull (W), Fréchet (F), and GEV survival function and return period we have:

$$x_W = \mu + \sigma(\ln\left(E(T)\right) - \ln\left(E(T) - 1\right))^{1/\beta}, \quad (3.4)$$

$$x_F = \mu + \frac{\sigma}{\left(\ln\left(E(T)\right) - \ln\left(E(T) - 1\right)\right)^{\beta}}, \quad (3.5)$$

$$x_{GEV} = \mu + \frac{\sigma}{\beta(\ln\left(E(T)\right) - \ln\left(E(T) - 1\right))^{\beta}} - \frac{\sigma}{\beta}. \quad (3.6)$$

Table 4. Periods of extreme occurence of weight of snow

Return period (year)	Frequency	Maximum annual weight of snow, kg/m^2			
		Gumbel	Weibull	Fréchet	GEV
2	$5\cdot10^{-1}$	68.2	51.2	62.4	62.3
5	$2\cdot10^{-1}$	125.1	100.9	106.6	106.5
10	10^{-1}	162.8	138.5	143.8	143.7
20	$5\cdot10^{-2}$	199.0	176.1	187.1	186.9
50	$2\cdot10^{-2}$	245.8	225.8	256.7	256.3
100	$\mathbf{10^{-2}}$	**280.9**	**263.4**	**321**	**320.5**
1 000	10^{-3}	396.8	388.3	639,9	638.3
10 000	10^{-4}	512.5	513.2	1219.9	1215.8
100 000	10^{-5}	628.2	638.1	2276.4	2266.7
1 000 000	10^{-6}	743.9	763	4200.6	4179

The maximum annual weight of snow appearance in various return periods is presented in Table 4. Every second year the maximum annual weight of snow is 68.2 kg/m^2 (Gumbel), 51.2 kg/m^2 (Weibull), 62.4 kg/m^2 (Fréchet), and 62.3 kg/m^2 (GEV). The expected maximum annual weight of snow over 100 years is 280.9 kg/m^2 (Gumbel), 263.4 kg/m^2 (Weibull), 321 kg/m^2 (Fréchet), and 320,5 kg/m^2 (GEV). There are no significant differences between maximum annual weight of snow when frequency is 10^{-2}; when frequency increases Fréchet and GEV maximum annual weight of snow values become much large then Gumbel and Weibull values. These latter differences are related to long tailes of these distributions. The return period of various extreme value distributions are shown in Fig. 10.

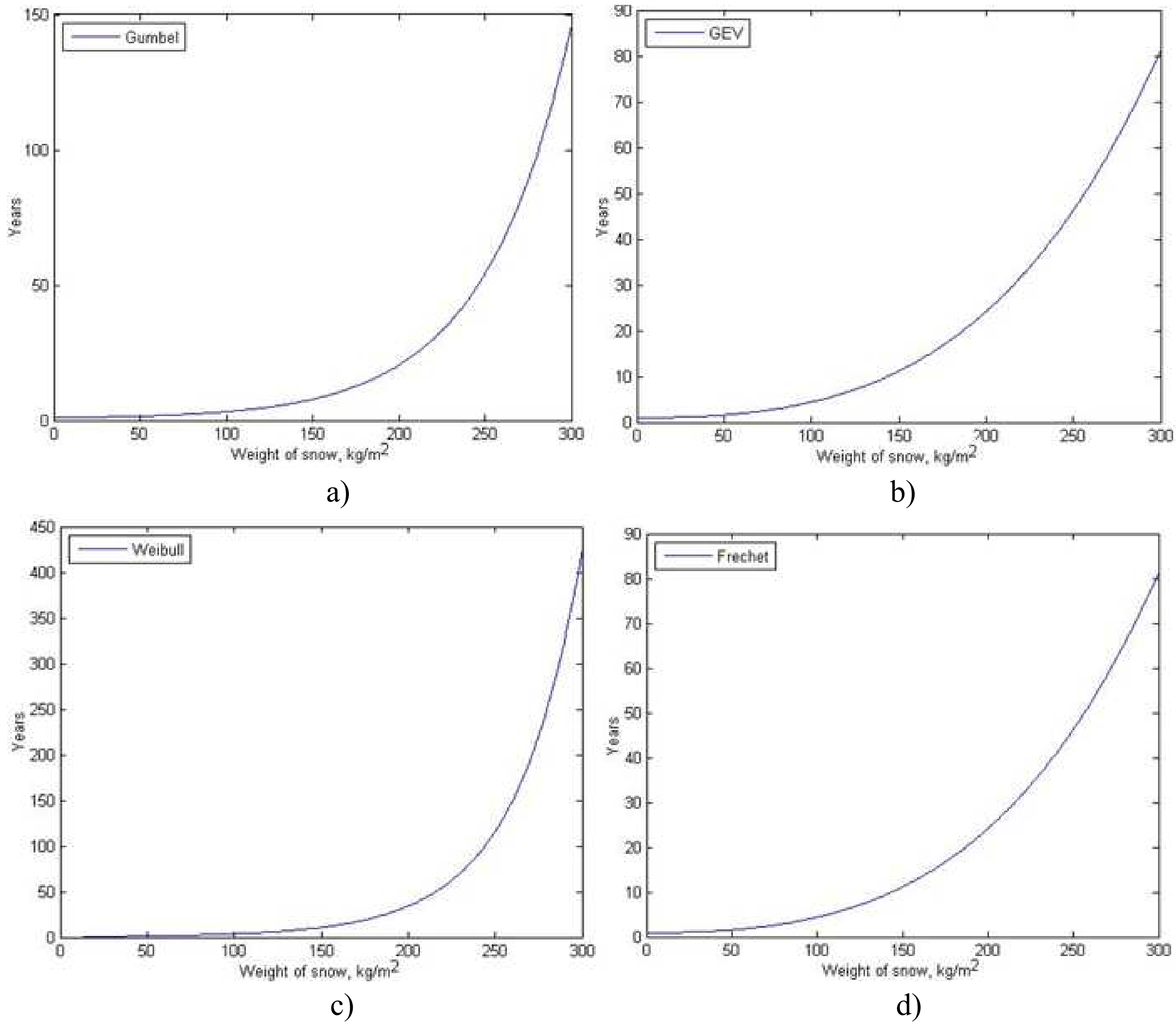


Fig. 10. Return period: a) Gumbel, b) GEV, c) Weibull, and d) Fréchet

The main disadvantage of graphical analysis is that it is difficult to evaluate the return period of extreme weight of snow (e.g., when x is: 50 kg/m^2, 100 kg/m^2, 200 kg/m^2 , and 300 kg/m^2). Table 5 shows estimated extreme weights of snow for various return periods.

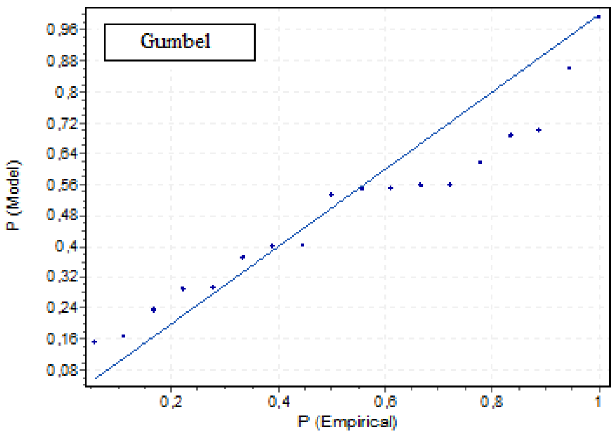
Table 5. Estimated weights of snow for different return periods

Gumbel				
Weight of snow, kg/m^2	50	100	200	300
Probability	$6.306 \cdot 10^{-1}$	$3.079 \cdot 10^{-1}$	$4.905 \cdot 10^{-2}$	$6.849 \cdot 10^{-3}$
Return period, year	2	3	20	146
Weibull				
Probability	$6.360 \cdot 10^{-1}$	$2.613 \cdot 10^{-1}$	$2.94 \cdot 10^{-2}$	$2.3 \cdot 10^{-3}$
Return period, year	2	4	34	434
Fréchet				
Probability	$6.421 \cdot 10^{-1}$	$2.283 \cdot 10^{-1}$	$4.16 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$
Return period, year	2	4	24	80
GEV				
Probability	$6.412 \cdot 10^{-1}$	$2.280 \cdot 10^{-1}$	$4.14 \cdot 10^{-2}$	$1.23 \cdot 10^{-2}$
Return period, year	2	4	46	81

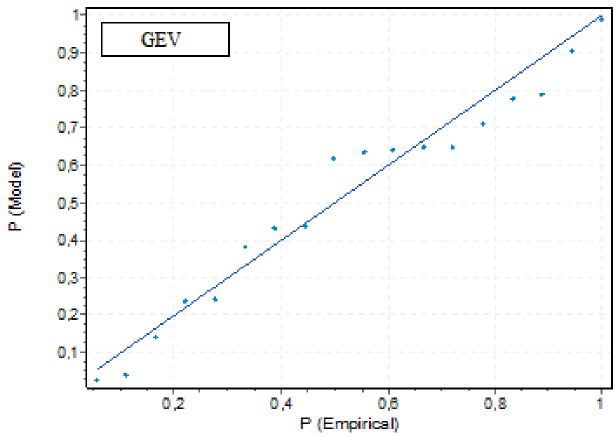
Return period for x equal to 50 kg/m^2 is two years for all extreme value distributions under consideration. This means that exceedance of weight of snow equal to 50 kg/m^2 is possible every second year. There are no considerable differences between return period of x equal to 100 kg/m^2 . Return period of 200 kg/m^2 is 20 years (Gumbel), 34 years (Weibull), 24 (Fréchet), and 46 (GEV). Actually, the relationship between any extreme weight of snow and return period can be determined by using presented models and statistics.

3.3 Optimal extreme value distribution

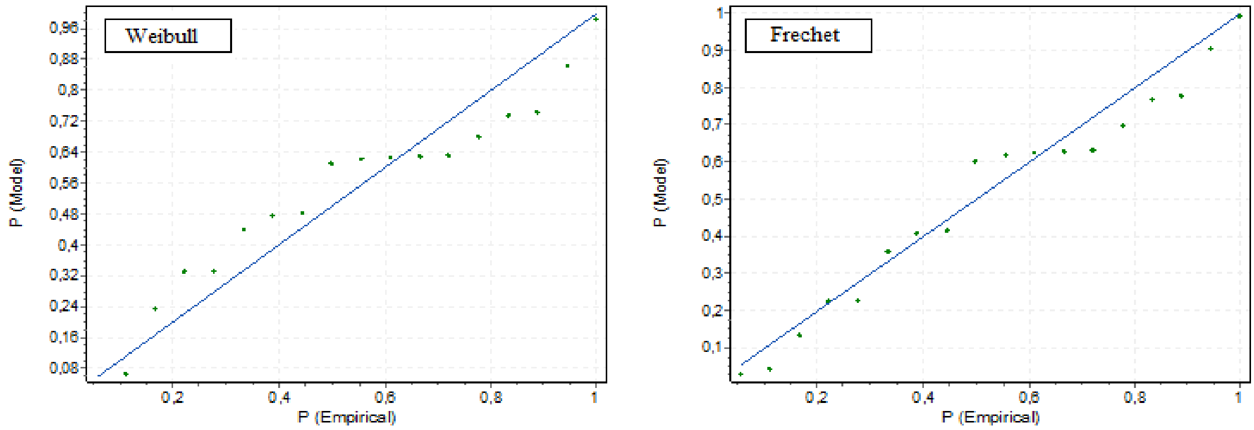
This section mainly presents graphical models and *goodness-of-fit* tests allowing to evaluate which of Extreme Value Distribution mentioned above has the best fit for the data.



a)



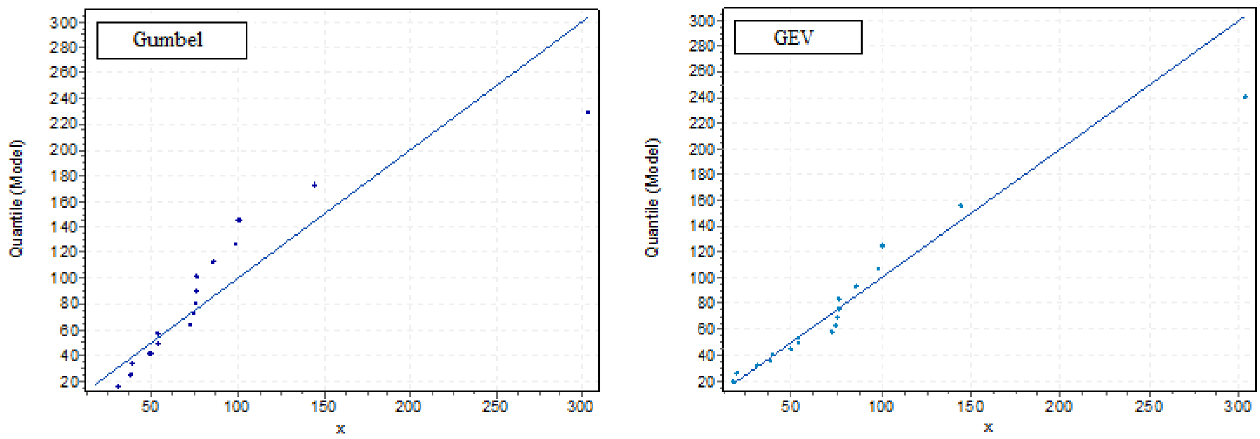
b)



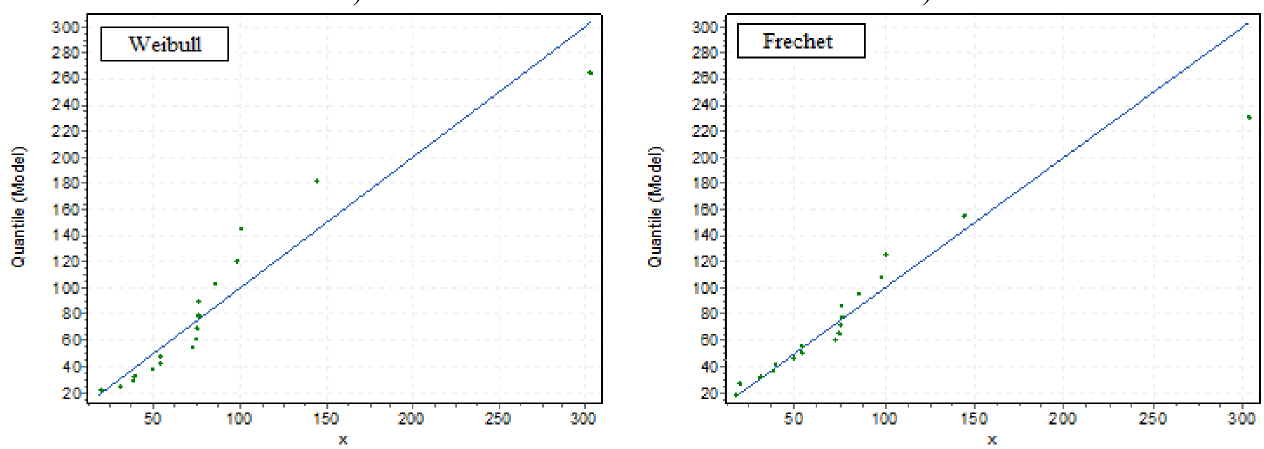
c) d)

Fig. 11. P–P plot: a) Gumbel, b) GEV, c) Weibull, and d) Fréchet

Fig. 11 presents P–P plots of extreme weight of snow according to all extreme value distributions shown above.



a) b)



c) d)

Fig. 12. Q–Q plot: a) Gumbel, b) GEV, c) Weibull, and d) Fréchet

Fig. 12 presents Q–Q plots of extreme weight of snow according to all extreme value distributions shown above. Both P–P and Q–Q plots shows reasonably good fitting of empirical and theoretical CDFs. More noticable differences between empirical and theoretical CDF can be seen from probability difference graphs shown below.

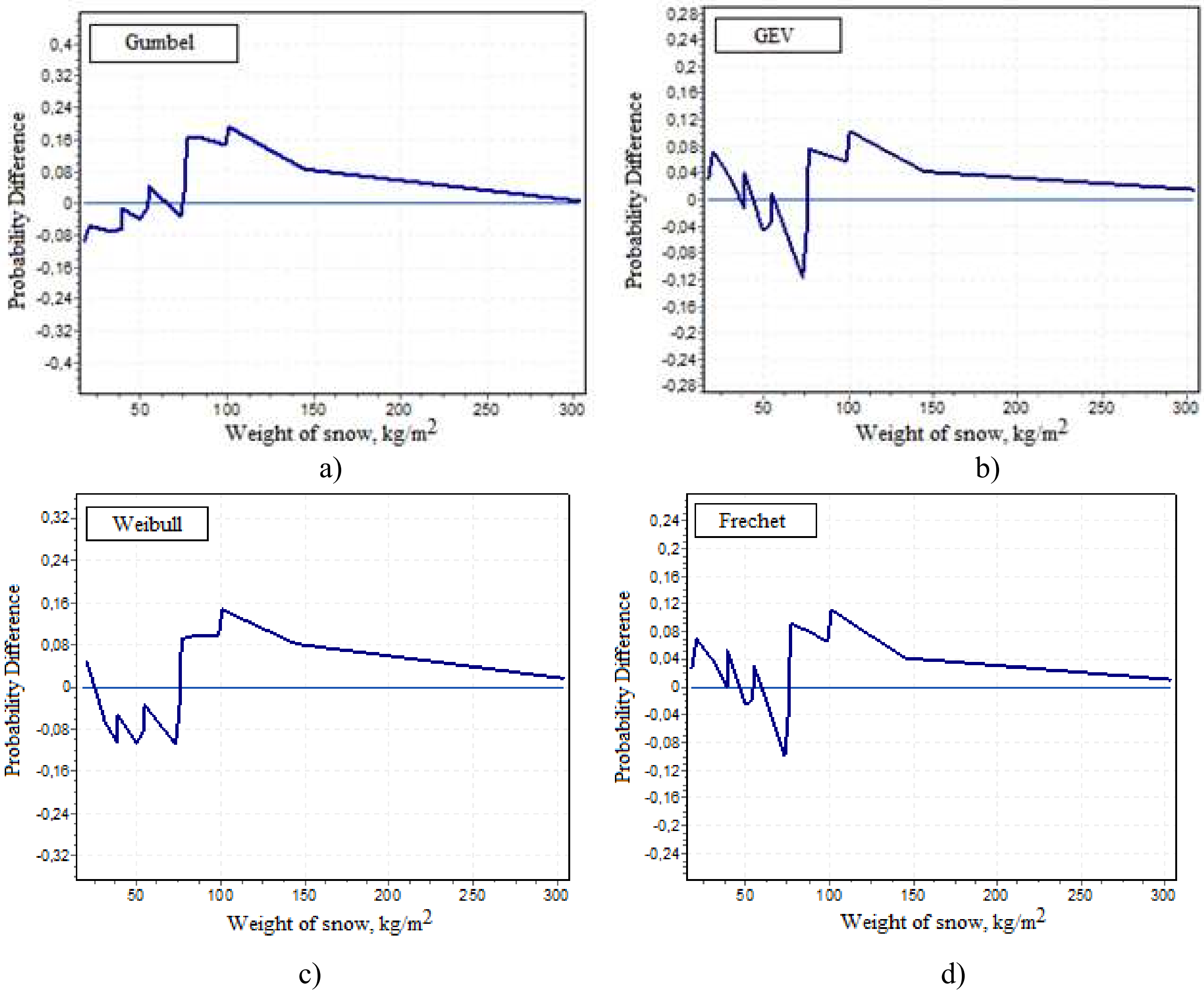


Fig. 13. Probability difference graph: a) Gumbel, b) GEV, c) Weibull, and d) Fréchet

Fig. 13 shows the probability differences between empirical and the theoretical CDF of four extreme value distributions: Gumbel, GEV, Weibull, and Fréchet. The bigest difference between empirical and theoretical CDF is 0.19 (Gumbel), 0.12 (GEV), 0.15 (Weibull), and 0.11 (Fréchet) when the weight of snow is: a) 100 kg/m^2, b) 75 kg/m^2, c) 100 kg/m^2, and d) 100 kg/m^2. A significant extreme weight of snow value with the biggest difference is 100 kg/m^2. Fréchet extreme value distribution has the smallest difference between empirical and theoretical CDF.

Grafical method has its advantages and disadvantages. The main advantage is that it is visually easy to compare several graphs at the same time, but it does not give exact values. In this case the *goodness-of-fit* (GOF) test is used to except the distribution that has the best fit to the data. *Godness-of-fit* test consists of Chi-Square, Kolmogorov-Smirnov, and Anderson-Darling tests.

Table 6. *Goodness-of-fit* tests estimation

Distribution	Chi-Square		Kolmogorov-Smirnov		Anderson-Darling
	p-Value	Statistic χ^2	p-Value	Statistic D	Statistic A^2
Gumbel	0.7095	0.19075	0.47197	0.19075	0.90727
Weibull	–	–	0.66365	0.16340	4.43900
GEV	0.51591	0.42207	0.60336	0.17177	0.36076
Fréchet	0.84502	0.03821	0.72934	0.15422	0.32083

Table 6 shows calculated GOF statistics and p-values for each of the fitted distributions.

Table 7. Acceptance of H_0 hypothesis at probability level $\alpha = 0.05$

Distribution	*Goodness-of-fit* tests		
	Chi-Square	Kolmogorov-Smirnov	Anderson-Darling
Gumbel	YES	YES	YES
Weibull	YES	YES	NO
GEV	YES	YES	YES
Fréchet	YES	YES	YES

Gumbel model's Chi-Square statistic χ^2 is 0.19075 with alpha level of significance equal to 0.05. The corresponding probability is $P < 0.2$. This is bigger than the conventionally accepted significance level of 0.05, so the null hypothesis that the two distributions are the same is accepted. Since χ^2 statistic (0.19075) do not exceeded the critical value for 0.05 probability level (0.30936) the null hypothesis can be accepted (YES). Table 7 shows the null hypothesis acceptance/rejection (YES/NO) for Gumbel, Weibull, GEV, and Fréchet distributions according to Chi-Square, Kolmogorov-Smirnov, and Anderson-Darling statistics at probability level $\alpha = 0.05$.

Graphical models and *goodness-of-fit* tests allowed us to except Fréchet distribution having the best fit for the data.

3.4 Uncertainty and sensitivity analysis

Since uncertainty analysis is performed with respect to Fréchet distribution location μ, scale σ, and shape β parameter estimates, a more comprehensive analysis is necessary for these three parameter samples: to find the distribution of samples and to evaluate main statistical characteristics. Due to insufficient amount of annual data more extreme weight of snow data are generated. 100 values are generated with Fréchet distribution parameters $\mu = -66.10$, $\sigma = 116.84$, and $\beta = 3.84$ in each year in order to find the new Fréchet distribution parameters of every year data set. In this case there are new data set of μ, σ, and β parameters. The next step is to find the best fitting distribution to these parameters. Table 8. Parameter estimates of probabilistic model shows the distribution of each parameter.

Table 8. Parameter estimates of probabilistic model

#	Parameters	Limits		σ	μ	Distribution
		Min.	Max.			
1	*shape* β	1.5801	4.8648	1.037	3.3895	Normal
2	*scale* σ	39.787	133.13	30.806	93.212	Normal
3	*location* μ	-81.978	4.8265	29.67	-45.367	Normal

All parameters are distributed by normal distribution with particular σ and μ values. Limits shows the maximum and minimum values of Fréchet distribution parameters (location, scale, and shape) data set. In further analysis 100 shape β (Par. 1), scale σ (Par. 2), and location μ (Par. 3) parameter values are generated according to normal distribution with σ, μ, minimum, and maximum values; also the extreme weight of snow probability is calculated in different weight *x*: 1, 10, 50, 100, 150,…, 300 kg/m^2. Consequently, 8 groups of 100 members are obtained. It follows the probabilistic curves that were evaluated at 100 different parameters values. 100 parameters pairs were chosen so that the uncertainty of the results can be evaluated with a 0.95 confidence level.

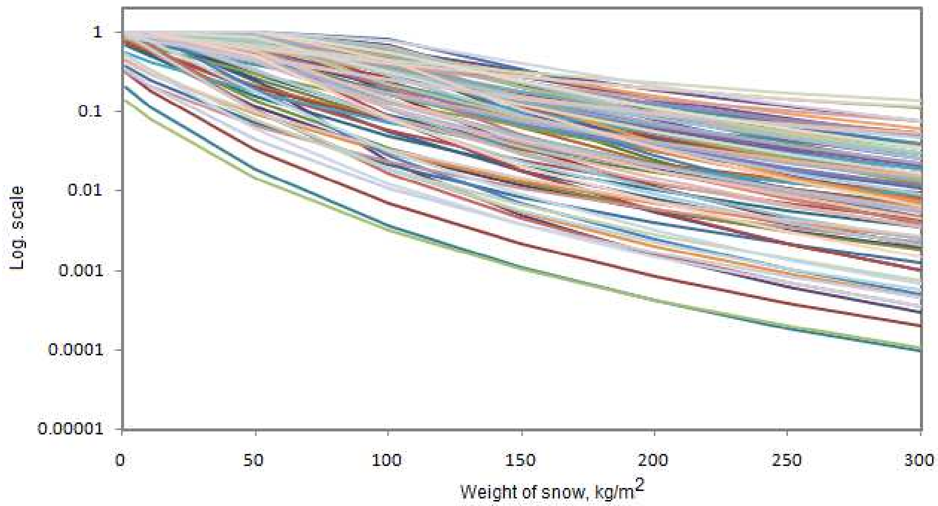
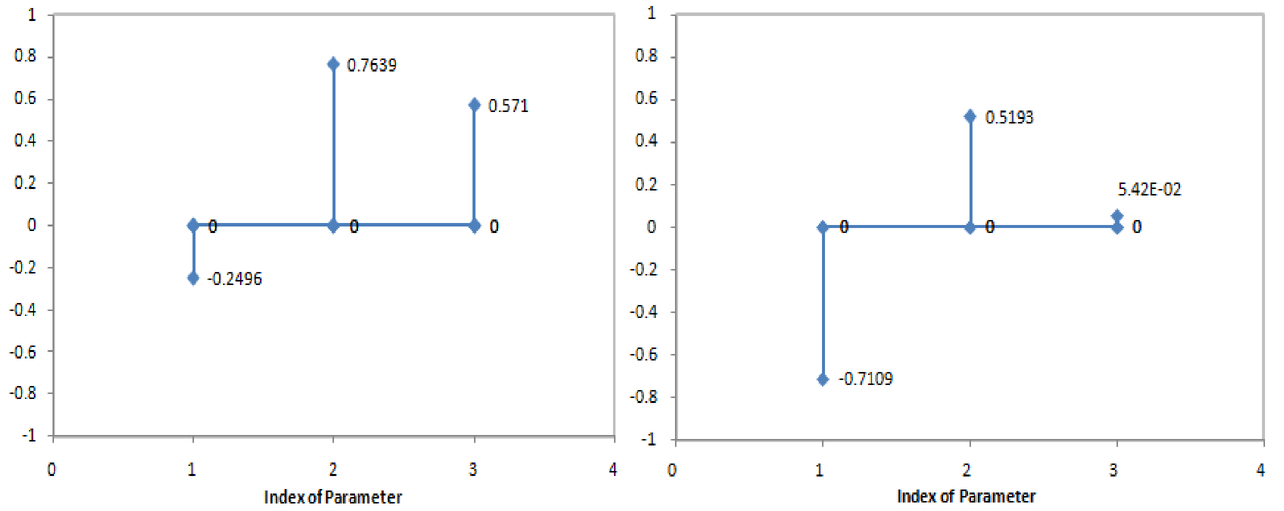


Fig. 14. Extreme weight of snow probabilistic curves

100 probabilistic curves visually shows the uncertainty limits of extreme weight of snow (Fig. 14). The probability estimates changes significantly depending on the weight of snow. It is clear that when small weights of snow is considered then interval of possible probabilities is wide and as weights of snow increases probability estimates decline rapidly.



a) b)

Fig. 15. Standardized regression coefficient of: a) 100 kg/m² and b) 300 kg/m² weight of snow

The sensitivity indicators are evaluated using statistical analysis in order to describe weight of the snow influence of the initial parameters and model parameters on results of a probabilistic model. Sensitivity indicators (standardized regression coefficients), that were obtained for weight of snow (Fig. 15): 100 kg/m² and 300 kg/m² shows, that for the first case results $(X \geq 100\,kg/m^2)$ mostly depends on scale σ parameter (Par. 2); for the second case $(X \geq 300\,kg/m^2)$ mostly depends on shape β parameter (Par. 1).

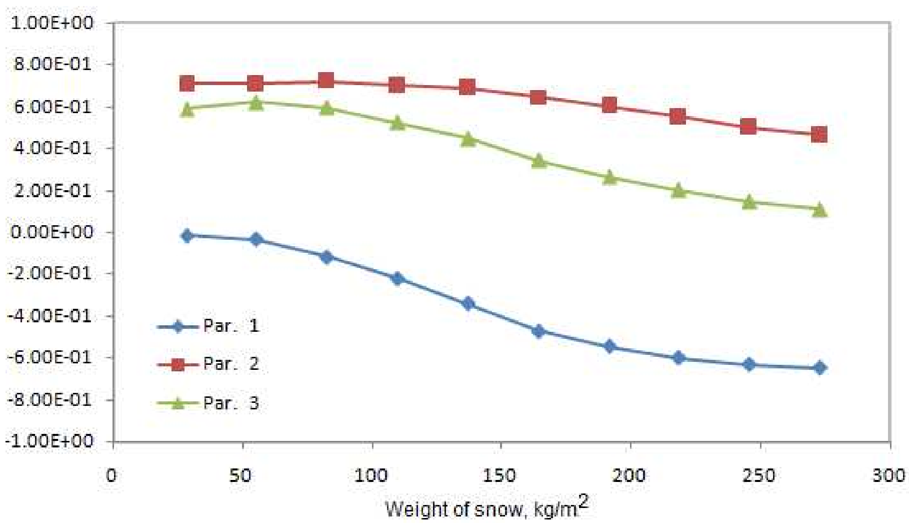


Fig. 16. Variation of standardized regression coefficient

Sensitivity analysis is best described by regression and correlation coefficients. Three model parameters (Par.1, Par. 2, Par. 3) as standardized regression coefficients change according to weight of snow x (Fig. 16). The model result are mostly influenced by scale σ parameter (Par. 2) and least influenced by shape β parameter (Par. 1) for all weights of snow.

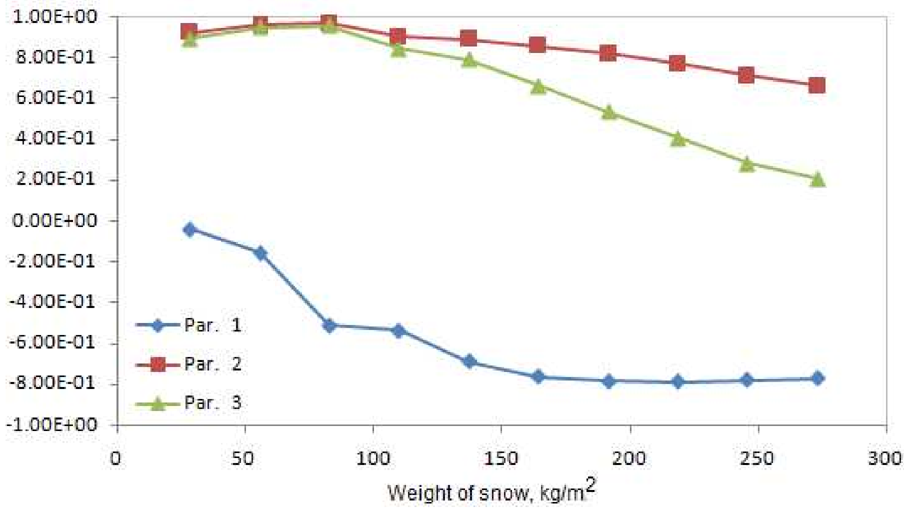


Fig. 17. Variation of partial correliation coefficient

Partial correlation analysis is useful because of its ability to evaluate the link between individual parameters and the model results. Calculated partial correlation coefficients supports conclusion about the influence of parameters on the model results.

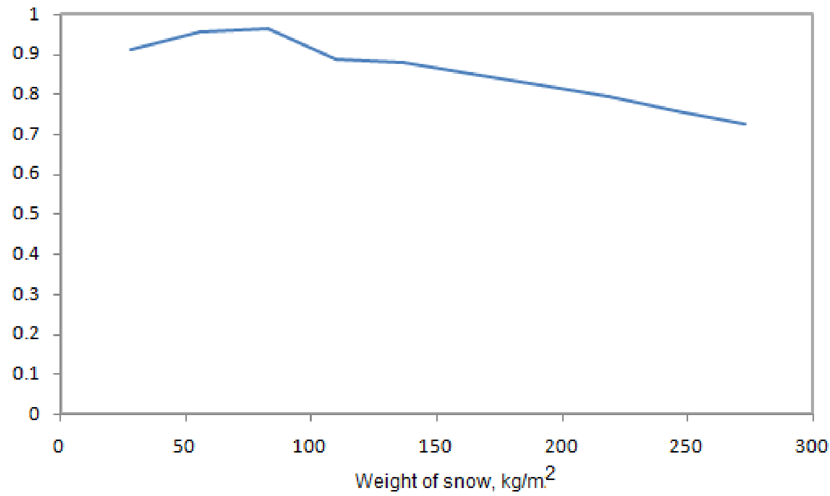


Fig. 18. Variation of coefficient of determination R^2

One of the most important factors is the coefficient of determination R^2. This coefficient shows what proportion of dispersion in the results can be explained by linear regression. Fig. 18 presents the change of coefficient of determination R^2 at various weights of snow x. Regression model is able to explain over 90 percent of a probabilistic model uncertainties when weight of snow x is less than 110 kg/m^2. However, when weight of snow exceeds 110 kg/m^2 level, uncertainty and sensitivity analysis results can be unreliable. This is because of the high dispersion of parameters and lack of information about these events.

CONCLUSIONS

1. The probabilistic assessment is based on mathematical models that were applied for statistical data of extreme snow weights in the period from 1992 up to 2009. Gumbel, Weibull, Fréchet, and GEV models were chosen for assessment of the likelihood of extreme weight of snow and allowed to evaluate the probability of different weights of snow x .
2. The return period of 50 kg/m^2 weight of snow is two years (for all distributions under consideration) according to the statistics and relationship between weight of snow and return period. Return period of 100 kg/m^2 weight of snow is 3 years (Gumbel), 4 years (Weibull, Fréchet, GEV) and of 300 kg/m^2 weight of snow is 146 years (Gumbel), 434 years (Weibull), 80 years (Fréchet), and 81 years (GEV).
3. Maximum annual weight of snow is calculated. The expected maximum annual weight of snow over 100 years is 280.9 kg/m^2 (Gumbel), 263.4 kg/m^2 (Weibull), 321 kg/m^2 (Fréchet), and 320.5 kg/m^2 (GEV).
4. Standardized regression and partial correlation analysis shows that the model result are mostly influenced by scale σ parameter (Par. 2) and least influenced by shape β parameter (Par. 1) for all weight of snow values.
5. Evaluated determination coefficient R^2 shows that regression model is able to explain over 90 percent of a probabilistic model uncertainties when weight of snow x is less than 110 kg/m^2. However, when weight of snow exceeds 110 kg/m^2 level, uncertainty and sensitivity analysis results can be unreliable.
6. This developed methodology may be used in the risk and safety analysis. For such analysis it is necessary to analyze the data in that area and recalculate probabilities of extreme weights of snow, also to assess the implications of the model and the uncertainty of results.

REFERENCES

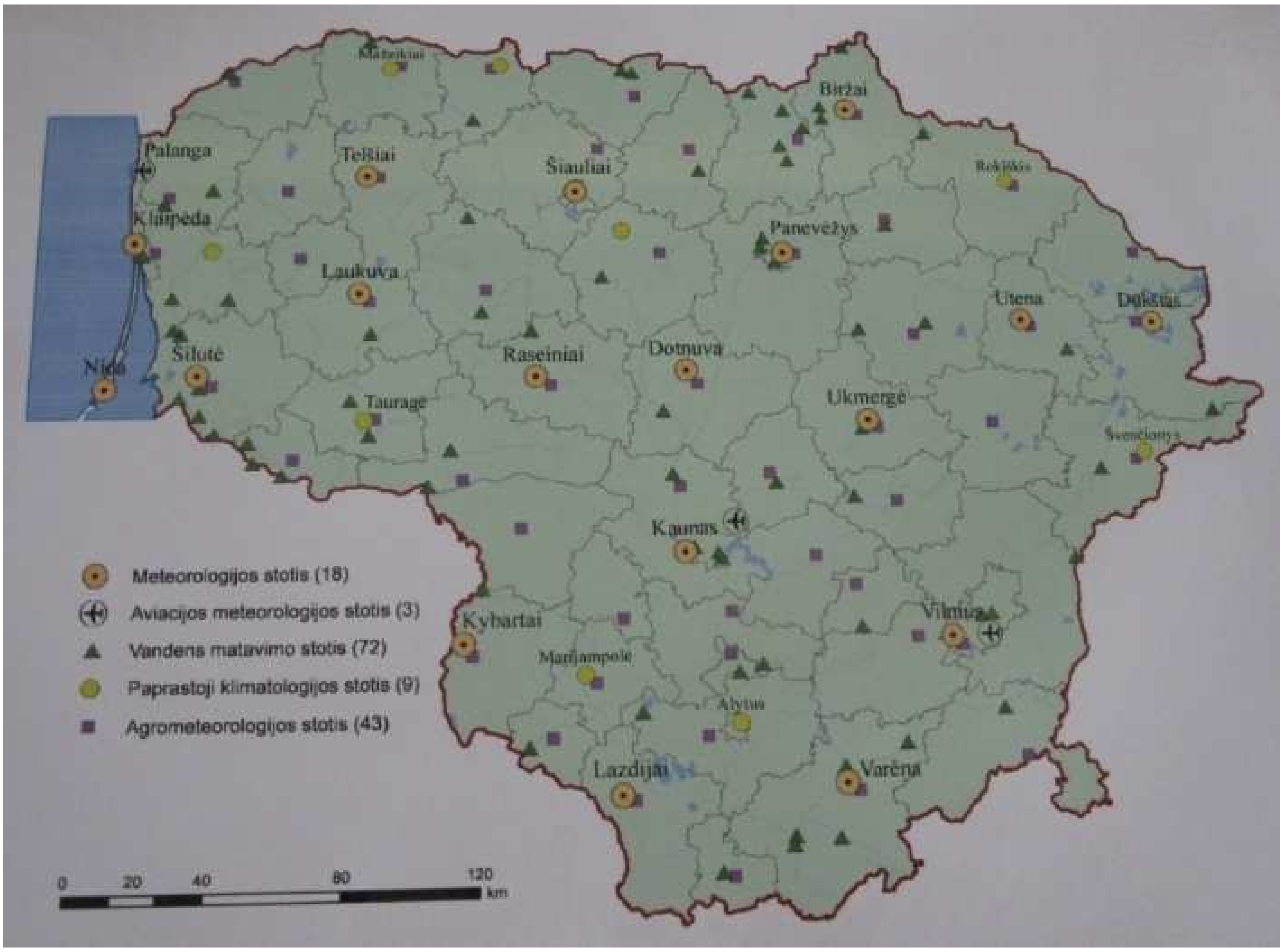
1. Aksomaitis, A. *Tikimybių teorija ir statistika.* Kunas: Technologija, 2002. 306 p. ISBN 9986-13-893-0.
2. Alzbutas, R.; Kupčiūnienė, K.; Adlytė, R.; Augutis, J. *Lėktuvo kritimo ant Ignalinos AE tikimybės vertinimas atsižvelgiant į duomenų neapibrėžtumą.* Energetika, 2007. Nr.1. 4 p.
3. Australian Government. Australian Bureau of Agriculcural and Recourse Economics and Science. Available at: http://adl.brs.gov.au/forestsaustralia/facts/fire.html 19/05/2011.
4. Australian Government. Available at: www.ausaid.gov.au/hottopics/topic.cfm?Id=9562_2054_7529_7688_4864 19/05/2011.
5. Beirlant, J.;Teugels, L.,J.; Vynckier, P. *Practical analysis of extreme values.* Leuven University Press, 1996. 24 p. ISBN 10 9061867681.
6. Castillo E. Extreme *Value Theory in Engenering*, London: Academic Press, Inc. Ltd. 1987. p. 5-9, p. 183-209.
7. Center for Satellite Based Crisis Information. Available at: www.zki.dlr.de/article/1348 19/05/2011.
8. Cox, D. ,R.; Oakes, D. *Analysis of Survival Data*. London: Chapman & Hall, 1984
9. Current Results. *Research news and Science facts*. Available at: http://www.currentresults.com/Weather-Extremes/Canada/precipitation.php 03/04/2011.
10. Embrechts, E.; Kluppelberg, C.; Mikosch, T. *Modelling Extremal Events for Insurance and Finance.* New York: Springer-Verlag, 1997. 152 p.
11. Farhangmehr, F. and Tumer, I. Y., *Capturing, Assessing, and Communication Tool for Uncertainty Simulation (CACTUS)*, Proceedings of ASME International Mechanical Engineering Congress & Exposition (IMECE 2008), Boston, MA.
12. Galambos, J.; Lechner, J.; Simiu, E. *Extreme Value Theory and Applications.* Proceedings of the Conference on Extreme Value Theory and Applications, Gaithersburg, Maryland, 1993, Vol.1. 16 p., 18 p. ISBN 0-7923-2865-5.
13. Glaeser, H. *GRS method for uncertainty and sensitivity evaluation of code results and applications.* Science and Technology of Nuclear Installations. 2008. Article ID 798901.7 p.
14. Gnanadesikan, R.; Wilk, M.B. *Probability plotting methods for the analysis of data. Biometrika* (Biometrika Trust), 1968. 55 (1). p. 1–17.
15. Gumbel, J.,E. *Statistics of Extremes.* New York: Dover Publications, Inc., 1958. 251 p. ISBN 0486-43604-7 (pbk.).

16. Yen, B.,C.; Cheng, S.,T.; Melching, C.,S. *First order reliability analysis: stochastic and risk analysis in hydroulic egineering.* In: Yen, B.C. (Ed.), Water Recourses Publications, Littleton, Colorado, 1986. p. 1-36.
17. Kambod A., H.; Mohammad R., M.; Mohammad K., B.; Masomeh R. *Engineering Geology and Geotechnical Aspects of Bam Earthquake.* Available at: http://www.iiees.ac.ir/iiees/English/bank/Bam/bam_report_english_geo.html 10/04/2011.
18. Kloos, M.; Hofer, E. *SUSA Version 3.2 User's guide and tutoria.* GRS, Garching, 1999.
19. Kopustinskas, V.; Alzbutas, R.; Augutis, J. *Matematinių modelių parametrų jautrumo ir rezultatų neapibrėžtumo statistiniai tyrimo metodai.* Energetika. 2007. T. 53. Nr. 3. p. 10-15.
20. Kopustinskas, V.; Alzbutas, R.; Augutis, J. *Matematinių modelių parametrų jautrumo ir rezultatų neapibrėžtumo statistiniai tyrimo metodai.* Energetika, 2007. T.53. Nr.3. p.12-14.
21. Kottegota, N.,T. *Stochastic Water Resources Technology.* Hong Kong: Mac. Press LTD, 1980.
22. Lietuvos Hidrometeorologijos Tarnyba prie Aplinkos ministerijos. *2006 metų stichinių hidrometeorologinių reiškinių apžvalga.* Vilnius. 2007. 3 p.
23. Lietuvos Hidrometeorologijos Tarnyba prie Aplinkos ministerijos. *2007 metų stichinių hidrometeorologinių reiškinių apžvalga.* Vilnius. 2008. 3 p.
24. Lietuvos Hidrometeorologijos Tarnyba prie Aplinkos ministerijos. *Hidrometeorologinės sąlygos 2008 metais.* Vilnius. 2009.
25. Lietuvos Hidrometeorologijos Tarnyba prie Aplinkos ministerijos. *Hidrometeorologinės sąlygos 2009 metais.* Vilnius. 2010.
26. Lietuvos Hidrometeorologijos Tarnyba prie Aplinkos ministerijos. Available at: http://www.meteo.lt/dokumentai/naujienos/20061019/pranesimas.ppt 12/04/2011.
27. *Meteorologinių stebėjimų nuostatai.* LHMT. 2004. Patvirtinti LHMT direktoriaus, 2004 m. balandžio 7 d. įsakymu Nr. V-51.
28. Ministry of Foreign Affairs of Japan. Available at: www.mofa.go.jp/j_info/visit/incidents/index.html 19/05/2011.
29. Murphy, P.N.D.; Xie, M.; Jiang, R. *Weibull Models.* New Jersey: John Willey & Sons, 2004, 9 p., 85 p., p. 86-87, p. 89-92. ISBN 0-471-36092-9.
30. Nelson, W. *Applied Life Data Analysis.* New York: Wiley, 1982.
31. O'Connor, D.,T.,P. *Practical Reliability Engineering.* London: Heyden & Son, 1995.
32. O'Connor, D.,T.,P. *Practical Reliability Engineering.* New York: Wiley, 1995.
33. Rachev, S.,T. *Handbook of heavy tailed distributions in finance.*(Handbook in finance). Elsevier. North Holand. 2003.

34. Rasimas, A.,B. Journal Energijos erdvė. 2011 Nr. 2(9). ISSN 2029-3119. 6 p.
35. Reiss, D., R.; Thomas, M. *Statistical Analysis of Extreme Values with Application to Insurance, Finance, hydrology, and Other Fields.* Berlin: Birkhauser-Verlag, 2007, 3rd ed. 61 p. ISBN 978-3-7643-7230-9.
36. Rinne, H. *The Weibull Distribution.* A Handbook. New York: Taylor & Francis Group, 2009. 317 p. ISBN-13: 978-1-4200-8743-7 (Hardcover).
37. Sachs, L. *Applied Statistcs.* Second edition. Springer Verlang, 1984.
38. Salvadori, G.; De Michele, C.; Kottegota, N. T.; Rosso, R. *Extremes in Nature. An Approach Using Copulas*. New York: Springer, Dordrecht, 2007, Vol. 56. ISBN 978-1-4020-4415-1.
39. Shiau, T.,J. *Return periodo f bivariate distributed extreme hydrological events.* Stochastic Environmental Research and Risk Assessment 17, Springler-Verlag, 2003. p. 42-57. DOI 10.2007/soo477-003-0125-9.
40. Stephens, M., A. *EDF Statistics for Goodness of Fit and Some Comparisons*. Journal of the American Statistical Association, 69, 1974. p. 730-737.
41. Stokhammar, P.; Oller, E.,L. *Density Forecasting of the Dow Jones Stock Index.* Research rapport. 2010. 14 p. ISBN 0280-7564.
42. Terminų žodynas. Available at: www.civilinesauga.lt/index.php?lng=lt&content=pages&page_id=18 26/05/2011.
43. Van Asselt, M.,B.,A. *Perspectives on uncertainty and risk: The PRIMA approach to decision support.* Dordrecht, The Netherlands: Kluwer Academic Publishers, 2000.
44. Walker, W.; Harremoes, P.; Rotmans, J.; van der Sluijs, J.; van Asselt, M.; Janssen, P & Krayer von Krauss, M. *Definding uncertainty: A conceptual basis for uncertainty management in model-based decision support.* Integrated Assessment 4(1), 2003. p. 5-18.
45. Wilks, S.,S. *Statistical prediction with special reference to the problem of tolerance limits.* Annals of Mathematical Statistics, Vol.13, 1942. p. 400-409.
46. Xu, P.Y.; Booij, J.,M.; Tong, B.,Y. *Uncertainty analysis in statistical modelis of extreme hydrological events.* Journal of the Stochastic Environmental Research and Risk Assessment. Springer – Verlag, 2009. DOI 10.1007/s00477-009-0337-8.

APPENDICES

Appendix 1. Meteorological stations in Lithuania

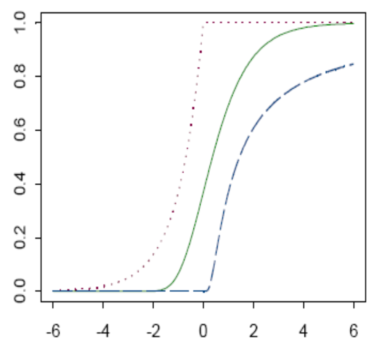
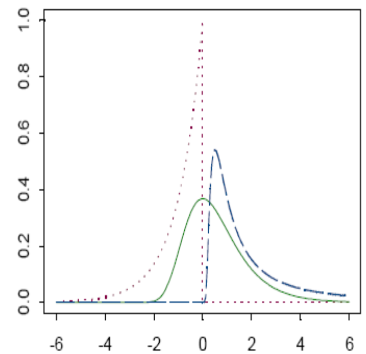


Appendix 2. Extreme weight of snow data

Year	Month	Snow thickness, cm	Snow density, g/cm^3	Snow water equivalent, mm	Weight of snow, g/cm^2
2009	1	16	0.11	11	1.76
	2	26	0.23	35	5.98
	3	27	0.27	35	**7.29**
	11				
	12	26	0.19	30	4.94
2008	1	11	0.14	11	1.54
	2	4			
	3	16	0.11	13	**1.76**
	11				
	12	4			
2007	1	15	0.06	5	0.9
	2	30	0.18	41	**5.4**
	3	12	0.34	27	4.08
	11	5	0.17	9	0.85
	12	4			
2006	1	12	0.23	18	2.76
	2	30	0.23	51	6.9
	3	30	0.25	55	**7.5**
	11				
	12	4			
2005	1	26	0.12	26	**3.12**
	2				
	3				
	11				
	12				
2004	1	44	0.21	67	9.24
	2	30	0.33	63	**9.9**
	3	25	0.3	45	7.5
	11	28	0.15	33	4.2
	12	5			
2003	1	30	0.1	24	3
	2	29	0.19	38	5.51
	3	33	0.23	51	**7.59**
	11				
	12	23	0.07	13	1.61
2002	1	46	0.22	66	**10.12**
	2	10	0.38	15	3.8
	3	7	0.16	10	1.12
	11	8	0.13	7	1.04
	12	23	0.1	14	2.3
2001	1	18	0.27	35	4.86
	2	28	0.21	36	5.88
	3	20	0.33	40	6.6
	11	9			
	12	45	0.17	51	**7.65**
2000	1	25	0.17	22	4.25
	2	14	0.39	12	**5.46**
	3	3			
	11				

	12	3			
	1	7	0.09	5	0.63
	2	43	0.2	58	**8.6**
1999	3	27	0.27	54	7.29
	11	8	0.1	6	0.8
	12	20	0.15	23	3
	1	20	0.1	9	2
	2	23	0.17	20	**3.91**
1998	3	20	0.17	24	3.4
	11	17	0.11	13	1.87
	12	21	0.11	19	2.31
	1	25	0.27	57	6.75
	2	10			
1997	3	1			
	11	13	0.12	12	1.56
	12	24	0.32	58	**7.68**
	1	44	0.18	56	7.92
	2	52	0.24	110	12.48
1996	3	66	0.22	117	**14.52**
	11	4			
	12	24	0.22	37	5.28
	1	15	0.26	26	**3.9**
	2	5			
1995	3	4			
	11	16	0.09	7	1.44
	12	21	0.13	20	2.73
	1	29	0.26		7.54
	2	44	0.69		**30.36**
1994	3	54	0.2		10.8
	11				
	12				
	1	16	0.2	18	3.2
	2	17	0.15	14	2.55
1993	3	25	0.2	26	**5**
	11	1			
	12	12	0.19	17	2.28
	1				
	2	17	0.12	13	**2.04**
1992	3				
	11	8			
	12				

Appendix 3. Extreme Value distributions

Distribution	Cumulative Distribution Function	Probability Density Function	Graphs of cumulative distribution function and probability density function
Gumbel	$F(x,\mu,\sigma)=e^{-e^{-\frac{x-\mu}{\sigma}}}, x \in R$	$f(x,\mu,\sigma)=\frac{1}{\sigma}e^{-(x-\mu)/\sigma}e^{-e^{-(x-\mu)/\sigma}}$	 
Fréchet	$F(x,\mu,\sigma,\beta)=\begin{cases}0 & x \le \mu \\ e^{-(x-\mu)/\sigma^{-\beta}} & x > \mu\end{cases}$	$f(x,\mu,\sigma,\beta)=\beta((x-\mu)/\sigma)^{(-\beta-1)}e^{-((x-\mu)/\sigma)^{-\beta}}$	
Weibull	$F(x,\mu,\sigma,\beta)=\begin{cases}e^{-(-(x-\mu)/\sigma)^{\beta}} & x < \mu \\ 1 & x \ge \mu\end{cases}$	$f(x,\mu,\sigma,\beta)=\frac{\beta}{\sigma}((x-\mu)/\sigma)^{\beta-1}e^{-((x-\mu)/\sigma)^{\beta}}$	

Appendix 4. Clasification of katastrophic events

	Ekstremalaus įvykio pavadinimas	Ekstremalaus įvykio apibūdinimas	Kriterijai	
			apibūdinimas (matavimo vienetas)	įvertinimas (dydis, reikšmė)
	I. GAMTINIO POBŪDŽIO			
1.	Geologinis reiškinys:			
1.1.	žemės drebėjimas	žemės drebėjimų virpesiai ir (ar) seisminis gruntų praskydimas, įvertinamas pagal Tarptautinę makroseisminę skalę MSK-64 ir Europos makroseisminę skalę EMS-98 [1]:		
		miesto teritorijoje	balai	≥5
		kitoje šalies teritorijoje	balai	≥7
		valstybės įmonėje Ignalinos atominėje elektrinėje (toliau vadinama – IAE), kitame branduolinės energetikos objekte	balai	≥7
1.2.	nuošliauža	natūralaus ar supilto šlaito nuošliauža, kelianti pavojų ypatingam objektui ar infrastruktūrai, – nuošliaužos paviršiaus plotas; poslinkio greitis	m^2; m per parą	≥100; ≥0,5
		Klaipėdos valstybinio jūrų uosto įplaukos ir laivybos kanalo povandeninio šlaito nuošliauža – šlaito poslinkio greitis pagal geodinaminio monitoringo duomenis	m per parą	≥1
1.3.	sufozinis reiškinys	sufozinės deformacijos trukmė; grunto sufozinės išnašos intensyvumas pažeistoje vietoje	para; m^3 per parą	≥1; ≥5
		Kauno hidroelektrinės, Kaišiadorių hidroakumuliacinės elektrinės, kito hidrotechnikos statinio grunto sufozijos sukelta deformacija ir (ar) sufozinė grunto išnaša	m^2	≥100
2.	Hidrometeorologinis reiškinys:			
2.1.	stichinis meteorologinis reiškinys:			
2.1.1.	labai smarki audra, viesulas, škvalas	maksimalus vėjo greitis	m/s	28÷32

	Ekstremalaus įvykio pavadinimas	Ekstremalaus įvykio apibūdinimas	Kriterijai	
			apibūdinimas (matavimo vienetas)	įvertinimas (dydis, reikšmė)
2.1.2.	smarkus lietus	kritulių kiekis per 12 valandų ir trumpiau	mm	50÷80
2.1.3.	ilgai trunkantis smarkus lietus	kritulių, iškritusių per 5 paras ir trumpiau, kiekis viršija vidutinį daugiametį mėnesio kritulių kiekį	kartai	2÷3
2.1.4.	stambi kruša	ledėkų skersmuo	mm	≥20
2.1.5.	smarkus snygis	kritulių kiekis; sniego dangos storis; trukmė	mm; cm; val.	20÷30; 20÷30; ≤12
2.1.6.	smarki pūga	vidutinis vėjo greitis; trukmė	m/s; val.	15÷20; ≥12
2.1.7.	smarki lijundra	apšalo storis ant standartinio lijundros stovo laidų	skersmuo, mm	≥20
2.1.8.	smarkus sudėtinis apšalas	apšalo storis ant standartinio lijundros stovo laidų	skersmuo, mm	≥35
2.1.9.	šlapio sniego apdraba	apšalo storis ant standartinio lijundros stovo laidų	skersmuo, mm	≥35
2.1.10.	speigas	nakties minimali temperatūra; trukmė	^{0}C; naktis	minus 30 arba žemesnė; 1÷3
2.1.11.	tirštas rūkas	trukmė; matomumas	val.; m	≥12; ≤100
2.1.12.	šalna aktyviosios augalų vegetacijos laikotarpiu	paros vidutinė oro temperatūra; oro (dirvos paviršiaus) temperatūra	^{0}C; ^{0}C	≤10; <0
2.1.13.	kaitra	dienos maksimali oro temperatūra; trukmė	^{0}C; diena	≥30; ≥10
2.1.14.	sausra aktyviosios augalų vegetacijos laikotarpiu	drėgmės atsargos dirvos sluoksnyje (0÷20 cm ir 0÷100 cm); hidroterminis koeficientas; trukmė	mm; skaitinė reikšmė; mėnuo	≤10 ir ≤60; <0,5; >1
2.2.	katastrofinis meteorologinis reiškinys:			
2.2.1.	uraganas	maksimalus vėjo greitis; tarptautinio pavadinimo suteikimas	m/s; suteikiamas uragano tarptautinis pavadinimas	≥33; taip
2.2.2.	labai smarkus lietus	kritulių kiekis; trukmė	mm; val.	>80; ≤12

	Ekstremalaus įvykio pavadinimas	Ekstremalaus įvykio apibūdinimas	Kriterijai	
			apibūdinimas (matavimo vienetas)	įvertinimas (dydis, reikšmė)
2.2.3.	ilgai trunkantis labai smarkus lietus	kritulių, iškritusių per 5 paras ir trumpiau, kiekis viršija vidutinį daugiametį mėnesio kritulių kiekį	kartai	>3
2.2.4.	labai smarkus snygis	kritulių kiekis; sniego dangos storis; trukmė	mm; cm; val.	>30; >30; <12
2.2.5.	labai smarki pūga	vidutinis vėjo greitis; trukmė	m/s; para	>20; ≥1
2.2.6.	smarkus speigas	nakties minimali temperatūra; trukmė	^{0}C; naktis	minus 30 arba žemesnė; >3

Appendix 5. MATLAB programme code

```
% GUMBEL MODEL

plot(data(:,1),data(:,2),'o','LineWidth',1,'MarkerEdgeColor','k','MarkerFaceColor','blac
k');
legend('Weight of snow, kg/m^2');
xlabel('year'); ylabel('kg/m^2');
grid off;
vid=mean(data(:,2))  % Mean
stdev=std(data(:,2))  %  Standard deviation
sigma=sqrt(6)*stdev/pi  %  Sigma
miu=vid-0.5772*sigma    % Miu
x=0:1:300;
Tikimybe=1-exp(-exp(-(x-miu)/sigma))  % Calculated Survival function values

figure(2);  % Gumbel survival function
plot(x,Tikimybe,'LineWidth',1);
legend('Gumbel survival function');
xlabel('Weight of snow, kg/m^2'); ylabel('Probability');

Tperiod=1./(Tikimybe);  % Return period
figure(3);
plot(x,Tperiod,'LineWidth',1);
legend('Gumbel Return period','location','northwest');
xlabel ('Weight of snow, kg/m^2'); ylabel ('Years');

%parmhat = wblfit(data(:,2))   % Weibull parameter estimates
%[parmhat,parmci] = wblfit(data(:,2))  % Returns 95% confidence intervals for the
estimates of a and b in the 2-by-2 matrix

% Gumbel and GEV survival functions
parmhat = gevfit(data(:,2)) %  GEV parameter estimates
P1 = gevcdf(x,-0.3033,29.7981,48.6465) % GEV cumulative distribution function
x=0:1:300;
figure(4)
plot(x,Tikimybe,x,1.-gevcdf(x,kMLE,sigmaMLE,muMLE),'LineWidth',1); %  Gumbel and GEV
Survival functions
legend('Gumbel','GEV');
xlabel('Weight of snow, kg/m^2'); ylabel('Probability');
title('Gumbel and GEV Survival functions')
text(23,0.94,' \leftarrow GEV','FontSize',10)
text(169,0.1,' \leftarrow Gumbel','FontSize',10)

% GEV MODEL

size(gevfit(data(:,2))) % returns the sizes of 'gevfit' dimension (3 values)
[paramEsts,paramCIs] = gevfit(data(:,2));  % returns 95% confidence intervals for the
parameter estimates.
kMLE = paramEsts(1)        % Shape parameter
sigmaMLE = paramEsts(2)    % Scale parameter
muMLE = paramEsts(3)       % Location parameter
kCI = paramCIs(:,1)        % 95% confidence interval for 'k'
sigmaCI = paramCIs(:,2)    % 95% confidence interval for 'sigma'
muCI = paramCIs(:,3)       % 95% confidence interval for 'miu'

% plot(parmhat,parmci)
[F,yi] = ecdf(data(:,2));   % Empirical cumulative distribution function
[nll,acov] = gevlike(paramEsts,data(:,2)); % Parameter standard errors
paramSEs = sqrt(diag(acov)) % Approximation to the asymptotic covariance matrix of the
parameter estimates
lowerBnd = muMLE-sigmaMLE./kMLE;  % The fitted distribution has zero probability below a
lower bound

% Density pdf plot
```

```
ymax = 1.1*max(data(:,2));
[F,yi] = ecdf(data(:,2));

% GEV survival function
x=0:1:300;
figure(5)
plot(x,1.-gevcdf(x,kMLE,sigmaMLE,muMLE),'-','LineWidth',1);
xlabel('Weight of snow, kg/m^2'); ylabel('Probability');
legend('GEV survival function','location','northeast');

% GEV return period
figure(6)
plot(x,1./(1.-gevcdf(x,kMLE,sigmaMLE,muMLE)),'-','LineWidth',1)
legend('GEV return period','location','northwest');
xlabel ('Weight of snow, kg/m^2'); ylabel ('Years');

% WEIBULL MODEL
y=13.6365:1:300;
sigma=68.3426;
beta=1.2568;
miu=13.6365;
TikWeib=exp(-(-(miu-y)/sigma).^beta) % Weibull survival function

figure(8)
plot(y,TikWeib) % Weibull survival function
figure(9)
plot(y,1./(TikWeib),'-','LineWidth',1) % Weibull return period
legend('Weibull','location','northwest');
xlabel ('Weight of snow, kg/m^2'); ylabel ('Years');


% plot the empirical cumulative distribution function of the data, fitting
% distribution according ML, LS methods
n = 18; % set of data
x=data(:,2);
%x = wblrnd(2,1,n,1);
x = sort(data(:,2));
p = ((1:n)-0.5)' ./ n;
stairs(x,p,'k-');
xlabel('x');
ylabel('Cumulative probability (p)');

logx = log(x);
logy = log(-log(1 - p));
poly = polyfit(logy,logx,1);
paramHat = [exp(poly(2)) 1/poly(1)]

plot(logx,logy,'+', log(paramHat(1)) + logy/paramHat(2),logy,'r--');
xlabel('log(x)');
ylabel('log(-log(1-p))');

paramMLE = wblfit(x)
stairs(x,p,'k');
hold on
xgrid = linspace(0,1.1*max(x),18)';
plot(xgrid,wblcdf(xgrid,paramHat(1),paramHat(2)),'r--', ...
     xgrid,wblcdf(xgrid,paramMLE(1),paramMLE(2)),'b--');
hold off
xlabel('x'); ylabel('Cumulative Probability (p)');
legend({'Data','LS Fit','ML Fit'},'location','southeast');

% Survival function of Gumbel model, Weibull using least squares line (LS),
% Weibull using maximum likelihood (ML) method
figure(2)
% plotting Weibull survival function using LS and ML methods
plot(xgrid,1.-wblcdf(xgrid,paramHat(1),paramHat(2)),'r--', ...
```

```
     xgrid,1.-wblcdf(xgrid,paramMLE(1),paramMLE(2)),'b--');

 vid=mean(data(:,2))       % Mean
stdev=std(data(:,2))       % Standard deviation
sigma=sqrt(6)*stdev/pi     % Sigma
miu=vid-0.5772*sigma       % Miu
x=0:0.1:300;
Tikimybe=1-exp(-exp(-(x-miu)/sigma))  % Gumbel survival function
hold on;
plot(x,Tikimybe,'LineWidth',1);  % plotting the Gumbel survival function
hold on;
xlabel('x'); ylabel('Survival function');
legend({'Weibull LS Fit','Weibull ML Fit','Gumbel'},'location','southeast');

% Finding Weibull 3 parameters (threshold method)
% we fix threshold parameter (miu=1) and find scale and shape parameters
% using least square method
x = sort(data(:,2));
p = ((1:n)-0.5)' ./ n;
logy = log(-log(1-p));
logxm1 = log(x-1);
poly1 = polyfit(log(-log(1-p)),log(x-1),1);
paramHat1 = [exp(poly1(2)) 1/poly1(1)]
figure(3)
plot(logxm1,logy,'b+', log(paramHat1(1)) + logy/paramHat1(2),logy,'r--');
xlabel('log(x-1)');
ylabel('log(-log(1-p))');
% Fix threshold parameter again (miu=2)
logxm2 = log(x-2);
poly2 = polyfit(log(-log(1-p)),log(x-2),1);
paramHat2 = [exp(poly2(2)) 1/poly2(1)]
% Fix threshold parameter again (miu=4)
logxm4 = log(x-4);
poly4 = polyfit(log(-log(1-p)),log(x-4),1);
paramHat4 = [exp(poly4(2)) 1/poly4(1)]

% Compare the lines with miu=1, miu=2, miu=4
figure(4)
plot(logxm1,logy,'b+', logxm2,logy,'r+', logxm4,logy,'g+', ...
     log(paramHat1(1)) + logy/paramHat1(2),logy,'b--', ...
     log(paramHat2(1)) + logy/paramHat2(2),logy,'r--', ...
     log(paramHat4(1)) + logy/paramHat4(2),logy,'g--');
xlabel('log(x - c)');
ylabel('log(-log(1 - p))');
legend({'Threshold = 1' 'Threshold = 2' 'Threshold = 4'}, 'location','northwest');

% We maximise the R^2 value over all possible threshold values in order to
% estimate the threshold parameter
r2 = @(x,y) 1 - norm(y - polyval(polyfit(x,y,1),x)).^2 / norm(y - mean(y)).^2;
threshObj = @(c) -r2(log(-log(1-p)),log(x-c));
cHat = fminbnd(threshObj,.75*min(x), .9999*min(x));
poly = polyfit(log(-log(1-p)),log(x-cHat),1);
paramHat = [exp(poly(2)) 1/poly(1) cHat]
logx = log(x-cHat);
logy = log(-log(1-p));
figure(5)
plot(logx,logy,'b+', log(paramHat(1)) + logy/paramHat(2),logy,'r--');
xlabel('log(x - cHat)');
ylabel('log(-log(1 - p))');

% Plot of Weibull CDF

sigma=68.3426;
beta=1.2568;
miu=13.6365;
```

```
i = 1;

for x=1:0.1:miu,
    Tik(i) = 1;
    iksai(i) = x;
    i = i + 1;
end;

for x=miu:300,
    Tik(i)=exp(-((x-miu)/sigma).^beta);
    iksai(i) = x;
    i = i + 1;
end;

figure(7);
plot(iksai, Tik);
xlabel('Weight of snow, kg/m^2'); ylabel('Probability');
legend('Weibull survival function','location','northeast
```