



Mallinnusta ja tulvien ennustamista

Ari Koistinen

Tutkija, Suomen ympäristökeskus

Matematiikan tuntiopettaja, Helsingin ammattikorkeakoulu Stadia

Mitä on mallinnus?

Mallinnusta on jo vuosikymmenien ajan käytetty erilaisten luonnossa, tekniikassa, talouselämässä ja yhteiskunnassa esiintyvien asioiden ja ilmiöiden tutkimiseen. Tietokoneiden suorituskyvyn kasvun myötä mallinnuksen sovellusmahdollisuudet ovat lisääntyneet valtavasti.

Mitä tämä mallinnus on? Hyvin laajasti ymmärrettynä asian kuvaaminen millä tahansa toisella asialla – siis toisen asian käyttäminen *mallina* – on mallinnusta. Esimerkiksi kartta on karttaa vastaavan alueen malli. Myös historian kokeeseen valmistautuvan lukiolaisen muodostama käsitys toisen maailmansodan tapahtumista on malli. Nämä molemmat mallit, kuten mallit useimmiten, ovat epätäydellisiä kuvauksia: kartassa ei voi olla kaikkia maaston todellisia yksityiskoh- tia ja lukiolaisen päähänsä rakentamasta mallista todennäköisesti puuttuu ainakin muutama maailmanso- dan tapahtumien sivujuonne.

Jos ilmiön tai muun asian kuvaamiseen käytetään matemaattisia yhtälöitä, on kysymys *matemaattisesta mallinnuksesta*. Yksinkertaisimmillaan matemaattinen malli on yksi ainoa yhtälö: esimerkiksi yhtälö $h = km$ kertoo, kuinka kokonaishinta h riippuu kilohinnasta k ja massasta m . Jos malli ohjelmoidaan tietokoneelle, saadaan *tietokonemalli*.

Edellä sanoja malli ja mallinnus käytettiin hyvin väljästi. Jokapäiväisessä tieteessä, tutkimuksessa ja tuotekehityksen kielenkäytössä mallinnuksella tarkoitetaan useimmiten ilmiön jäljittelemistä matemaattisella kuvauksella, joka on hieman monimutkaisempi kuin yksi ainoa yhtälö, ja usein kuvaukseen kuuluu peruslaskutoimitusten lisäksi monimutkaisempaa matematiikkaa. Kuvaus tai ainakin osia siitä on myös useimmiten ohjelmoitu tietokoneelle niin, että ilmiön kulkua eri tilanteissa voidaan tarkastella tietokoneen avulla. Tietokone kuuluu mallinnukseen usein niin olennaisena osana, että itse matemaattinen malli ja mallin kuvauksen sisältämä tietokoneohjelma mielletään samaksi asiaksi.

Mallin kehittäminen, esimerkkinä vesistömalli

Yritysten, tutkimuslaitosten ja korkeakoulujen tutkimus- ja tuotekehitystoiminnassa käytetään paljon valmiita eri tarkoituksiin tehtyjä mallinnusohjelmistoja. Valmiita työkaluja on tarjolla virtausmallinnukseen, rakenteiden ja koneen osien mallinnukseen, logistiikan simulointiin (esim. tehtaan tuotteiden ja raaka-aineiden kuljetus ja varastointi), elektronisten järjestelmien mallinnukseen – ja oikeastaan lähes jokaiselle insinööritieteen osa-alueelle. Tällaisen mallin etuna on usein helppokäyttöisyys: kuvitellaan

esimerkiksi elektronista piiriä kuvaava malli, johon syötetään tiedot toisiinsa kytketyistä komponenteista, minkä jälkeen malli kertoo, kuinka piiri todellisuudessa toimisi, näyttäen esimerkiksi piirin eri osissa kulkevat virrat sekä komponenttien jännitteet ja varaukset eri ajanhetkillä.

Toisaalta valmiit työkalut ovat usein pitkälle erikoistuneita ja samalla rajoittuneita: jotakin mallinnettavaan ilmiöön vaikuttavaa seikkaa ei ehkä olekaan mahdollista ottaa huomioon mallinnusohjelmassa. Esimerkiksi elektronisten piirien mallinnustyökalussa ei välttämättä voida huomioida ympäristön lämpötilan äkillisiä muutoksia ja niiden vaikutusta piirin komponentteihin. Tällöin koko ohjelmaa olisi muutettava, mutta lähdekoodi tuskin on vapaasti muokattavissa. Ohjelmistojen tarkkoja toimintaperiaatteita ei välttämättä liikesalaisuuden säilyttämiseksi paljasteta, mikä saattaa häiritä mallia käyttävää asiantuntijaa.

Edellä mainittujen syiden lisäksi myös valmiiden mallinnusohjelmistojen korkeat hinnat saattavat vaikuttaa siihen, että ilmiön mallintamiseksi päätetäänkin tehdä kokonaan uusi malli: laaditaan matemaattinen kuvaus ja sen pohjalta tietokonemalli jollain ohjelmointikielellä. Seuraavassa tarkastellaan pintapuolisesti joi-takin tällaisen mallin kehittämisen vaiheita, käyttäen esimerkkinä vesistöjen virtaamien ja vedenkorkeuksien ennustamiseksi laadittua mallia.

Suomen ympäristökeskuksessa (SYKE) on noin kahdenkymmenen vuoden ajan kehitetty ja käytetty vesistömallijärjestelmää, jolla simuloidaan ja ennustetaan vedenkorkeuksia ja virtaamia. Satojen ennustepisteiden laskentatulokset ovat päivittäin nähtävillä internetissä. Järjestelmän ytimenä olevan mallin syötteenä on sää (lämpötila, sadanta ja haihdunta), joka perustuu havaintoihin, sääennusteeseen tai tilastolliseen dataan. Malliin on ohjelmoitu karkea kuvaus veden kiertokulusta ilmakehässä, pinta- ja pohjavesissä sekä maaperässä. Pintavesien kuvaus on näistä yksityiskohtaisin ja mallin päätarkoituksena onkin simuloida järvien ja jokien vedenkorkeuksia ja virtaamia sekä eri vesistöalueilta valuvia vesimääriä.

Vesistömallin toimintaperiaate on perusidealtaan yksinkertainen. Tilat, joissa vesi luonnossa esiintyy, on kuvattu varastoina, joita ovat mm. järviaaltaat, maaperän vesivarastot (maan pintakerros ja pohjavesi) ja talvella lumipeite. Jokaista varastoa vastaa mallin jokaisella osa-alueella muuttuja, jonka arvo muuttuu laskennan edetessä. Mallin laskenta alkaa tietystä päivästä, jolloin kullekin muuttujalle annetaan alkuarvo. Alkuarvojen määrittämisessä käytetään hyödyksi myös mahdolliset havainnot. Tämän jälkeen edetään yleensä vuorokauden mittaisissa laskenta-askelissa ja siirretään vettä ohjelmakoodissa olevien hydrologian lainalaisuuksiin perustuvien yhtälöiden avulla varastosta (eli muuttujasta) toiseen.

Laskennan tulokset esitetään vedenkorkeutta, virtaamaa ja muita tärkeitä suureita ajan funktiona kuvaavina käyriä. Myös mahdolliset havainnot näkyvät kuvissa. Ennuste esitetetään todennäköisyysjakaumana, jonka hajonta syntyy pääasiassa sään epävarmuuden perusteella.

Vesistömallijärjestelmää käytetään mm. tulvien ennustamiseen, vesistöjen säännöstelyn suunnitteluun sekä vaikeasti suoraan mitattavissa olevien hydrologisten suureiden kuten valunnan (alueelta valuva vesimäärä millimetreinä vuorokaudessa) ja maankosteuden vaihtelun laskemiseen.

Ennen tietokonemallin laatimista on oltava siis ainakin jonkinlainen käsitys lainalaisuuksista, joita ilmiö noudattaa. Jos nämä lainalaisuudet tunnetaan hyvin tarkasti, niin riittää ohjelmoida niitä kuvaavat matemaattiset yhtälöt tietokoneelle sekä laatia käyttöliittymä mallin ohjaamista ja laskentatulosten esittämistä varten. Tämän jälkeen malli on valmis testattavaksi ja kun ohjelmointi- ja muut virheet – joita luultavasti on melkoinen joukko – on saatu korjattua ja malli todettu toimivaksi, se voidaan ottaa käyttöön.

Usein tieto mallinnettavaan ilmiöön liittyvistä lainalaisuuksista on kuitenkin enemmän tai vähemmän puutteellisia: vaikuttavia tekijöitä on niin paljon, että tarkkan matemaattisen kuvauksen laatiminen on mahdotonta. Puutteellinen tieto näistä lainalaisuuksista voidaan monissa tilanteissa korvata käyttämällä tuntemattomia parametreja, joille pyritään etsimään sopivat arvot.

Malleja on tapana jaotella moniin kategorioihin, kuten dynaamisiin ja staattisiin malleihin (dynaamisessa mallissa tapahtuu muutoksia ajan suhteen, staattisessa ei) sekä stokastisiin ja deterministisiin malleihin (stokastinen malli sisältää satunnaisuutta, mutta deterministisen mallin tulokset määräytyvät täysin syötetietojen perusteella). Eräs jaottelu perustuu siihen, kuinka tarkasti ilmiöt on mallissa kuvattu. *Fysikaalisessa mallissa* kuvaus perustuu tiukasti fysiikan lakeihin ja mahdollisten tuntemattomien parametrien arvot on voitu rajata suhteellisen pienelle alueelle: esimerkiksi mallinnettaessa nesteen virtausta paperikoneen sisällä tunnetaan ehkä ilmiötä kuvaavat osittaisdifferentiaaliyhtälöt ja paperikoneen rakenteiden geometriasta johtuvat reunaehdot, mutta ei välttämättä aivan tarkasti nesteen koostumusta ja sen fysikaalisia ominaisuuksia.

Tilastollisessa mallissa ei pyritä fysikaaliseen kuvaukseen, vaan tilastollista dataa analysoimalla yritetään selvittää, kuinka paljon jokin tarkasteltava asia riippuu eri tekijöistä – vai riippuuko lainkaan. Esimerkiksi talouden suhdanteiden kehittymistä voidaan pyrkiä ennustamaan tilastollisilla menetelmillä työllisyyden, vaihtotaseen, inflaation, yritysten investointien sekä kuluttajien taloudellisen luottamuksen perusteella, ilman, että

tarvitsee kuvata tarkasti prosessia, jonka kautta esimerkiksi kuluttajien luottamusindeksi vaikuttaa suhdanteisiin (toki on eduksi, jos mallintajalla on tästäkin asiasta jonkinlainen aavistus).

Konseptuaalinen malli tarkoittaa eräänlaista fysikaalisen ja tilastollisen mallin välimuotoa. Siinä prosessit on pyritty kuvaamaan todellisuutta vastaavalla tavalla, mutta kuvaus on karkea, ennemmin käsitteellinen kuin fysikaalinen. Tuntemattomilla parametreilla, joita voi olla paljon, on melko suuri sallittu arvoalue.

SYKEN vesistömalli on konseptuaalinen malli. Esimerkiksi vesistöalueelta toiselle valuva vesimäärä riippuu veden määrästä alueella sekä kyseisen alueen viivettä kuvaavista parametreista. Ei ole pyrittykään rakentamaan tarkkaa fysikaalista kuvausta veden virtaamisesta, sillä se edellyttäisi yksityiskohtaisia tietoja pinnanmuodoista, maaperän koostumuksesta ja kasvillisuudesta.

Veden virtauksen pois luonnontilaisista järviältäistä määrää havaintoihin perustuva purkauskäyrä (tai purkaustaulukko – järvestä purkautuva vesimäärä kuutiometreinä sekunnissa eri vedenkorkeuksille), ja jos virtaamamittauksia ei ole tai niitä on liian vähän kattavan purkaustaulukon muodostamiseksi, perustuu purkauskäyräkin parametreihin. Yleensä käytetään muotoa

$$q(w) = \begin{cases} a \cdot (w - w_0)^b, & \text{kun } w > w_0, \\ 0, & \text{muuten.} \end{cases}$$

olevaa funktiota, missä w on järven vedenkorkeus (metreinä merenpinnan tasosta luettuna) ja q on virtaama kuutiometreinä sekunnissa. Luvut a , b ja w_0 ovat parametreja, joiden arvot on pyritty määrittämään niin, että laskenta toimii havaintoihin nähden mahdollisimman hyvin. Parametrille w_0 on myös yksinkertainen fysikaalinen tulkinta: se on raja, jonka alapuolisilla vedenkorkeuksilla virtaama on nolla.

Säännöstelyjen järvien lähtövirtaama määritetään mallissa säännöstelyyn liittyvien määräysten ja tiedossa olevien juoksu- ja suunnitelmien perusteella. Lumi- ja järvien kertyminen riippuu sadannasta, lämpötilasta ja muutamasta parametreista. Parametrit vaikuttavat myös siihen, kuinka nopeasti vesi valuu maan pintakerroksesta pohjaveteen ja kuinka nopeasti pohjaviesivaraston vesi siirtyy eteenpäin vesistöalueelta toiselle ja järviin.

Mallin kalibrointi

Mallin parametrien arvojen tulisi siis olla sellaiset, että mallin laskentatulokset ovat mahdollisimman lähellä havaintoja. Optimaalisten parametrien arvojen etsimistä kutsutaan mallin *kalibroinniksi*. Hyvin yksinkertaisissa tapauksissa kalibrointi voidaan tehdä käsin: annetaan parametreille arvot ja testataan ne ajamalla

malli. Tämän jälkeen päätellään laskentatuloksia ja havaintoja vertaamalla, mihin suuntaan parametreja tulee muuttaa. Tätä jatketaan, kunnes saadaan riittävän hyviä tuloksia – tai kunnes tulokset eivät enää merkittävästi parane.

Usein malli on niin monimutkainen, että kalibrointi on parempi antaa tietokoneen tehtäväksi. Kalibrointi on monesti laskennallisesti raskas tehtävä. Mallinnuksessa, kuten monilla muillakin tietotekniikan käyttöalueilla, laskentaresurssit ovat tietotekniikan huimasta kehityksestä huolimatta aina niukat: tietokoneiden nopeuden ja muistikapasiteetin kasvaessa keksitään yhä monimutkaisempia mallinnettavia asioita tai kasvatetaan vanhojen mallien laskentatarkkuutta (esimerkiksi aika- tai paikkaresoluutiota). Nykyisin eräs laskennallisesti vaativimpia tietokoneiden työsaikoja lienee sääilmiöiden mallintaminen.

Mahdollisimman hyvään tulokseen pääsemiseksi sekä prosessoriajan (ja työajan) säästämiseksi mallin parametrien kalibrointiin tulisi käyttää tehtävään hyvin sopivaa menetelmää. Näitä *optimointimenetelmiä* matemaatikot ovat kehittäneet lukemattomia. Optimointi tarkoittaa funktion maksimi- tai minimikohdan etsimistä – etsitään siis sellaiset muuttujien arvot, että funktion arvo on mahdollisimman suuri tai mahdollisimman pieni. Mallin kalibroinnissa muuttujia ovat parametrit ja funktiona on mallin virhe, joka pyritään minimoimaan. Yksinkertainen ja yleisesti käytetty tapa muodostaa virhefunktio on *pienimmän neliösumman* periaate: lasketaan summa jokaisen havaintoarvon ja sitä vastaavan lasketun arvon erotuksen neliöistä, täsmällisesti ilmaistuna virhefunktio on

$$f(\bar{x}) = \sum_{k=1}^n (z_k(\bar{x}) - y_k)^2,$$

missä \bar{x} on mallin parametreista koostuva vektori, josta funktion arvo riippuu, luvut $z_1(\bar{x}), z_2(\bar{x}), \dots, z_n(\bar{x})$ ovat mallin laskemia (ja parametreista riippuvia) arvoja ja luvut y_1, y_2, \dots, y_n näitä vastaavat havaitut arvot. Lasketut ja havaitut arvot voivat olla esimerkiksi virtaamia vesistön eri kohdissa kuutiometreinä tietyn aikajakson, vaikkapa vuorokauden aikana. Jos funktion f arvo saadaan pieneksi, ovat lasketut ja havaitut arvot lähellä toisiaan.

Optimointimenetelmän on oltava sellainen, että virhefunktion minimi löydetään suhteellisen nopeasti ja riittävän suurella varmuudella. Tyypillinen optimoinnin ongelmatilanne on juuttuminen ns. paikalliseen minimiin: löydetään kohta, jossa funktion arvo on lähitöllä olevissa pisteissä saavutettuihin arvoihin nähden pieni. Mikäli kyseessä on yhden tai kahden muuttujan funktio (eli \bar{x} on reaali- tai kompleksivektorinen vektori), voidaan asia ilmaista havainnollisesti sanomalla, että funktion kuvaajassa on tässä kohdassa ”kuoppa”. Kauempana voi kuitenkin olla syvem-

piä kuoppia ja hyvän optimointimenetelmän tulisi melko erehtymättömästi löytää se syvin ja mielellään kohtuullisella prosessoriajalla.

Optimointiongelman ratkaisemistakin vaikeampaa on usein itse ongelman asettaminen, tarkemmin sanottuna virhefunktion muodostaminen. Jos pyritään laskemaan arvoja useille eri suureille, voi käydä niin, että parametrien arvojen muuttaminen saattaa parantaa yhtä, mutta huonontaa toista laskennan osa-alueita. Muodostettaessa virhefunktiota näille osatekijöille tulee valita tarkoituksenmukaiset painokertoimet. Esimerkiksi mallinnettaessa vesistöjä pitäisi lumen vesiarvon talven aikana vastata alueella tehtyjä lumimittauksia, mutta lisäksi kevään virtaamasumman tulisi olla lähellä havaittua virtaamasummaa, eikä olisi pahitteeksi, jos kevätvirtaaman huipun suuruus ja ajankohta osattaisiin ennustaa kohtuullisella tarkkuudella oikein.

Kun parametrit on optimoitu, ei vielä voida ottaa mallia operatiiviseen käyttöön tai laittaa sitä myyntiin. Edessä on mallin testaus, jonka jälkeen mallin rakennetta täytyy mahdollisesti – tai todennäköisesti – vielä muuttaa ja optimoida parametrit uudelleen. Tämä ei välttämättä pääty edes mallin varsinaiseen käyttöönottoon. SYKEN vesistömallin kehitystyö jatkuu edelleen, sillä vedenkorkeuksia ja virtaamia ei vielä pystytä ennustamaan riittävän tarkasti pitkällä eikä aina lyhyelläkään aikavälillä. Lähemmäksi tätä tavoitetta päästään myös sääennustemallien kehittyessä: sääilmiöt kun vaikuttavat vesistöjen käyttäytymiseen melko voimakkaasti.

Kuinka mallintajaksi tullaan?

Teknisen tai luonnontieteellisen – ja mahdollisesti myös taloustieteellisen, lääketieteellisen tai yhteis-

kuntatieteellisen – koulutuksen valinnut joutuu (tai pitäisi ehkä sanoa ”pääsee”) yhä suuremmalla todennäköisyydellä tekemisiin mallinnuksen kanssa, joko mallien kehittäjänä tai niiden käyttäjänä. Edellytykset suunnitella ja kehittää malleja ovat hyvät, jos mallinnettavien ilmiöiden tuntemisen lisäksi tuntee matematiikkaa ja osaa ohjelmoida – eivätkä nämä taidot ole pahitteeksi myöskään valmiita mallinnusohjelmistoja käytettäessä, sillä monet niistä edellyttävät käyttäjältään hyviä matemaattisia ja tietoteknisiä valmiuksia. Tarvitaan ”malliajattelua”, kykyä hahmottaa ilmiöitä niin, että ymmärtää syy- ja seuraussuhteet ja osaa puukea ne täsmälliseen ja tarvittaessa matemaattiseen muotoon.

Mallintajaksi voi päätyä opiskelemalla ensisijaisesti jotakin sovellusaluetta, jolla mallinnusta voi hyödyntää, ja hankkia sen ohella hyvät tiedot matematiikasta ja tietojenkäsittelystä. Toinen vaihtoehto on opiskella pääaineena matematiikkaa ja perehtyä lisäksi sovellusalueisiin, kuten tekniikan eri osa-alueisiin, fysiikkaan, kansantaloustieteeseen tai geofysiikkaan – tai melkein mihin hyvänsä, sillä on mahdotonta sanoa, mihin kaikkeen matemaattista mallinnusta tulevaisuudessa sovelletaan.

Lopuksi pari aiheeseen liittyvää linkkiä:

Matemaattisen mallinnuksen verkostohanke (täältä löytyvästä linkkikokoelmasta voi aloittaa lisätietojen etsimisen mallinnuksesta):

<http://alpha.cc.tut.fi/mallinnus/>

SYKEN vesistöennusteet:

<http://www.ymparisto.fi/vesistoennusteet>