



**XXXII. Országos
Tudományos Diákköri
Konferencia**

2014



**COSMO-RS momentumok alkalmazása
hőmérsékletfüggő folyadék-gáz határfelületi
feszültség becsléséhez**

Készítette:

Kontos János

Témavezetők:

Dr. Dallos András

Dr. Járvás Gábor

Kémia Intézet, Fizikai Kémia Intézeti Tanszék
Mérnöki Kar

Pannon Egyetem
Veszprém

Tartalom

Összefoglaló.....	1
1. Bevezetés	2
2. Irodalmi összefoglaló.....	3
2.1 A gáz-folyadék határfelületi feszültség	3
2.2 COSMO-RS elmélet és a σ -momentumok	5
3. Az általam fejlesztett becslőmódszer bemutatása.....	11
3.1 A bemeneti adatbázis létrehozása	11
3.2 Többváltozós lineáris regresszió.....	13
3.3 Mesterséges neurális hálózatok	15
4. Eredmények	21
4.1 Többváltozós lineáris regresszió.....	22
4.2 Mesterséges neurális hálózatok	25
5. Összefoglalás	30
6. Irodalomjegyzék	32
7. Köszönetnyilvánítás.....	34

Összefoglaló

A 0,1 μm -nél kisebb átmérőjű cseppek párolgása esetén - a csepp görbült geometriája miatt - a párolgó anyagok gőznyomását a Kelvin-egyenlet szerint kell számítani. A Kelvin-összefüggés a csepp átmérője mellett a komponensek folyadék-gáz határfelületi feszültség értékeit is figyelembe veszi. Mivel a csepp párolgása nem izoterm folyamat, ezért figyelembe kell vennünk a felületi feszültség hőmérsékletfüggését is. Ezért egy olyan módszer, mely csupán a molekulaszervezet felhasználásával képes az adott molekulára jellemző illesztendő paraméterek nélkül, a felületi feszültség becslésére különböző hőmérsékleten, nagy érdeklődésére tarthat számot. Ezért célul tűztam ki egy olyan becslő módszer kifejlesztését, mely a molekulák felületi töltés-sűrűség eloszlását jellemző paraméterek (COSMO-RS σ -momentumok mint molekuláris jellemzők) és a hőmérséklet felhasználásával képes a felületi feszültség értékének pontos becslésére.

A rendelkezésemre álló szakirodalom kritikus feldolgozásával létrehoztam egy adatbázist, mely 187 vegyület, különböző hőmérsékleten kísérletileg meghatározott folyadék-gáz határfelületi feszültség értékeit tartalmazza. Ezt követően kvantumkémiai COSMO számítás segítségével meghatároztam az egyes anyagok σ -momentumait. Elsőként többváltozós lineáris regressziót (MLR) végeztem, vagyis a σ -momentumok és a felületi feszültség között lineáris kapcsolatot felételeztem. Mivel sokváltozós minimumkeresési feladatról van szó, ezért többféle minimumkeresési eljárást is megvizsgáltam. A MLR segítségével nem sikerült megfelelő korrelációt elérnem, ezért a mesterséges neurális hálózatot (ANN) - mint nem-lineáris problémák leírására alkalmas fekete doboz modell - használtam. A mesterséges neurális hálózat optimalítása során vizsgáltam a neuronok számának, a tanítási algoritmusnak, illetve a neuron átviteli függvényének hatását a becslés minőségére.

A legjobbnak talált „idegháló” esetén a tanítási és teszt szetthez 0,967, illetve 0,961 négyzetes korrelációs együttható értékek tartoznak, ami a mért és a számított felületi feszültség értékek jó egyezését igazolta. A kidolgozott modell vizsgálatával lehetőségem nyílt a felületi feszültség létrehozásában fontos szerepet játszó intermolekuláris erők hatásának értelmezésére is.

A kutatás a TÁMOP 4.2.4.A/2-11-1-2012-0001 azonosító számú „Nemzet Kiválóság Program - Hazai hallgatói, illetve kutatói személyi támogatást biztosító rendszer kidolgozása és működtetése konvergencia program” című kiemelt projekt keretében zajlott. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Kulcsszavak: COSMO-RS, gáz-folyadék határfelületi feszültség, mesterséges neurális hálózat (ANN)

1. Bevezetés

Napjainkban a különböző tulajdonságbecslő módszerek számítógépes alkalmazása szerves részét képezik a mérnöki gyakorlatnak. A termodinamikai tulajdonságok pontos meghatározása elengedhetetlen a különféle vegyipari berendezések tervezése során. A legfejlettebb tervezőszoftverek (Aspen Plus, PRO II, UniSim) már előre definiált tulajdonságbecslő modullal rendelkeznek, melyek a legtöbb esetben igen jól használhatóak. A felhasználónak nem kell ismernie a becslőeljárás pontos menetét, csupán ki kell választani egy komponenst a gyártó által feltöltött adatbázisból. Sajnos előfordulhat olyan eset, amikor egy olyan vegyülettel szeretnénk dolgozni, melyet, ezek az adatbázisok nem tartalmaznak. Ekkor saját magunknak kell elvégezni a tulajdonságbecslést. A különböző hőmérsékletfüggő termodinamikai tulajdonságok becslése esetén további problémát jelent az adott tulajdonság hőmérsékletfüggésének leírása.

Munkám során egy olyan becslőmódszert hoztam létre, mely képes a molekulák szerkezetét felhasználva a hőmérsékletfüggő gáz-folyadék határfelületi feszültség becslésre. A kifejlesztett becslőmódszer jól illeszkedik az eddigi munkám során fejlesztett többkomponensű cseppek párolgását leíró modellbe. A kisméretű ($0,1\ \mu\text{m}$ kisebb) átmérőjű cseppek esetén az anyagok gőznyomása megnő a csepp görbült geometriája miatt. Ennek a változásának a leírására a Kelvin-egyenlet használható. A Kelvin-egyenletben szereplő gáz-folyadék határfelületi feszültség így jelentősen befolyásolja az anyagok gőznyomás értékeit kisméretű cseppek esetén, ezáltal meghatározza a párolgás teljes folyamatát.

A kifejlesztett becslőmódszer más területeken is jól alkalmazható. A különböző folyadékok heterogén katalizátorként való használata során a szilárd hordozó és a folyadék közötti kapcsolat kialakításának tervezésekor elengedhetetlen a felületi feszültségek pontos ismerete. A vegyipari berendezések (rektifikáló kolonnák, extrakciós oszlopok) tervezésekor és méretezésekor nagy szükség van a különféle folyadékok gáz-folyadék határfelületi értékeinek meghatározására.

A becslőmódszer létrehozása során törekedtem a legújabb elméleti eszközök (COSMO-RS elmélet, mesterséges neurális hálózatok) alkalmazására. Ezek az eszközök lehetőséget adtak arra, hogy a becslőmódszer alkalmazhatóságának spektruma a lehető legszélesebb lehessen.

2. Irodalmi összefoglaló

Az irodalmi összefoglalóban bemutatásra kerülnek a gáz-folyadék felületi feszültséghez és annak becsléséhez kapcsolódó elméletek és módszerek. Az összefoglaló második részében a kifejlesztett becslőmódszer alapját képező COSMO-RS elmélet és a σ -momentumok ismertetésére kerül sor.

2.1 A gáz-folyadék határfelületi feszültség

A gáz-folyadék határfelületi feszültség (γ) a folyadékok jellemző tulajdonsága, mely hatására a folyadék minden esetben igyekszik a minimális felszint biztosító formát felvenni. Ennek oka, hogy a folyadék felszínén elhelyezkedő molekulákra ható erők eredője a folyadékfázis belseje felé mutat. A felületi feszültség definíció szerint az az energia mely ahhoz szükséges, hogy a folyadék felületét egységnyivel megváltoztassuk [1]. A folyadékok felületi feszültsége a hőmérséklet növekedésével csökken. A változás leírására az Eötvös Lórándról elnevezett Eötvös-szabály használható. A gáz-folyadék határfelülethez kapcsolódó jelenségek száma igen nagy. Többek között ide tartozik a lótosz-effektus és a különféle kapilláris jelenségek is. A felületi feszültség jelentőségét jól mutatja, hogy 2012-ben sikerült elsőként egy vízcseppet kettévágni [2] hidrofób késsel.

A gőz-folyadék határfelületi feszültség kísérleti meghatározására számos módszer alkalmazható. A mérési eljárások egy része a felületi feszültséggel kapcsolatban álló jelenségre (kapilláris emelkedésre és süllyedésre, a folyadékcsepp alakjának változása) van visszavezetve, míg más módszerek közvetlenül képesek mérni a felületi feszültséget (kiszakításos módszer). A gáz-folyadék határfelületi feszültség fontosságát mutatja, hogy jelenleg is folynak fejlesztések az újabb mérési eljárás kidolgozására [3]. A mérési módszerek nagy számával ellentétben, a gáz-folyadék határfelületi feszültség becslésére csupán korlátozott számú eljárás áll rendelkezésünkre [4]. Ping Le és munkatársai [5] által közölt csoportadalékos becslőmódszer a megfelelő állapotok tételén alapszik. A módszer telített, telítetlen, gyűrűs és aromás illetve oxigént tartalmazó szénhidrogének esetén alkalmazható. Hátránya, hogy a becslésre szolgáló összefüggések bemenő paraméterei között szerepel a kísérletileg meghatározott normális forráspont. Abbas és Nordholm [6] munkájukban egy általánosított van der Waals elméleten alapuló becslőmódszert mutatnak be. Sajnos csak egy vegyületcsalád (hidrogén halogenidek) esetén alkalmazható az eljárás. Romero-Martinez és munkatársai [7] által létrehozott

modell egyenes láncú és elágazásos láncú szénhidrogének hőmérsékletfüggő gáz-folyadék határfelületi feszültségének becslésére használható. A módszer bemenő paraméterei között szerepel az adott szénatom számhoz tartozó alkán felületi feszültségének értéke, mely csökkenti a modell széleskörű alkalmazhatóságát.

A mérnöki gyakorlatban legelterjedtebb gáz-folyadék határfelületi feszültség becslésére szolgáló eljárás a MacLeod-Sugden módszer és a Grain módszer. Az eredetileg MacLeod [8] által kifejlesztett empirikus becslőmódszer a gáz-folyadék határfelületi feszültség és a folyadékfázis sűrűsége között teremt kapcsolatot. Később Sugden [9] rámutatott arra, hogy az a MacLeod által felállított empirikus egyenletben szereplő konstansokat összevonva megkaphatjuk az ún. parachort értéket. A parachor egy adott anyagra jellemző paraméter, kapcsolatot teremt a folyadék felületi feszültsége, sűrűsége és molekulatömege között. A Grain módszer [4] hasonló a Ping Le és munkatársai által közölt módszerhez, ugyanis ebben az esetben is szükség van kísérletileg meghatározott bemeneti paraméterre (normális forráspont). Látható, hogy a gáz-folyadék határfelületi feszültség becslésére szolgáló becslőmódszerek kis száma mellett sokszor problémát jelent az adott anyagra jellemző, kísérletileg meghatározandó paraméterek hiánya.

Az általam létrehozott becslőmódszer az adott molekulákra jellemző σ -momentumok felhasználásával képes a hőmérsékletfüggő gőz-folyadék határfelületi feszültség becslésére. Azokat a regressziós modelleket, melyek egy molekuláris és egy makroszkópikus tulajdonság között teremtenek kapcsolatot ún. QSPR (Quantitative structure–property relationship) modelleknek nevezzük. Gardas és munkatársai [10] szobahőmérsékletű ionos folyadékok gáz-folyadék határfelületi feszültségének becslésére hoztak létre lineáris QSPR modellt. A módszer bemenő paraméterei az adott anyagra jellemző parachor értékek. A módszer széleskörű alkalmazhatóságának határt szab, hogy a parachor számítása során szükségünk van az ionos-folyadék adott hőmérsékleten kísérletileg meghatározott sűrűségére. A Kajeh és munkatársai [11] által létrehozott nem-lineáris QSPR modell, a szerzők által definiált molekuláris tulajdonságok és különböző szerves hűtőközegek gáz-folyadék határfelületi feszültsége között teremt kapcsolatot fuzzy modellek felhasználásával.

Látható, hogy egy olyan becslőmódszer, mely pusztán molekuláris jellemzőket felhasználva képes hőmérsékletfüggő gőz-folyadék határfelületi feszültség becslésére nagy érdeklődésére tarthat számot.

2.2 COSMO-RS elmélet és a σ -momentumok

A különböző makroszkópikus jellemzők (ide tartozik a gáz-folyadék határfelületi feszültség is) becslése mikroszkopikus, vagyis molekuláris tulajdonságok felhasználásával igen összetett. Az elemek (molekulák) egyéni tulajdonságai mellett szükségünk van a közöttük kialakuló kapcsolatok modelljére is. Ezen információk birtokában vagyunk képesek a molekuláris jellemzőket egy mérhető, szabad szemmel is jól látható jelenség leírására felhasználni. Többféle megközelítés lehetséges, melyek közül a legfontosabbak a következőkben bemutatásra kerülnek.

A molekulák tisztán kvantumkémiai leírása szerint 0 K hőmérsékleten a részecske teljesen izolált a környezetétől [12]. A módszer hátránya, hogy a molekulák közötti kapcsolatokat nem tudja figyelembe venni. Jól használható ideális gázfázisbeli viselkedés meghatározására.

Az empirikus vagy félempirikus módszerek közé tartoznak az ún. csoportadalékos módszerek. Ekkor a molekulát felbontjuk kisebb csoportokra és az egyes csoportokhoz tartozó értékeket egy előzőleg meghatározott összefüggésbe helyettesítjük be, mely segítségével megkapjuk a kívánt jellemző becslt értékét. Egyszerűbb molekulák esetén (pl. alkánok, alkének) igen könnyen és gyorsan használható módszerek, melyeknek becslési pontossága szűkebb vegyületcsoportok esetén igen nagy lehet. Hátránya, hogy a bonyolultabb atomcsoportok esetén nem mindig áll rendelkezésünkre az adalék értéke, így ezekben az esetekben ezt a módszert sajnos nem tudjuk a csoportok közötti kölcsönhatásokat figyelembe vételére alkalmazni.

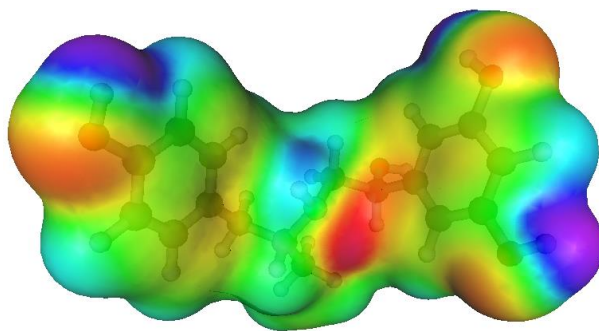
A mérnöki gyakorlatban sokszor előforduló folyadékelegyek, illetve oldatok fizikai-kémiai tulajdonságainak számítása még összetettebb feladat. Az ún. explicit modellezési technikák közül legszélesebb körben a Monte Carlo szimuláció és a Molekula Dinamika (MD) alkalmazható. Ezen folyamatok során adott számú részecske mozgását és az elemek között kialakuló kölcsönhatásokat követjük figyelemmel. Az eredményt valamilyen termodinamikai átlagolás segítségével kaphatjuk meg. Mivel az egyes molekulák mozgását külön-külön számítjuk, így a módszer számításigénye igen nagy. Az eredményt jelentősen befolyásolja a molekulák leírásának módja (pl. merev töltött gömb).

A Molekula Dinamika és a Monte Carlo szimulációval ellentétben léteznek implicit megközelítések melyek nem a molekulák közötti kölcsönhatásokat, hanem a molekula

és egy folytonos tömbfázis között fellépő elektrosztatikus hatásokat veszik figyelembe. Legegyszerűbben úgy lehet elképzelni, hogy a molekulát egy - a folytonos közegben kialakított - üregbe helyezzük. Ez az eset a valóságban nem fordulhat elő, viszont nagyon jól képes visszaadni azt az állapotot, amikor a molekulát számos más molekula veszi körül. Ezeket a megközelítéseket dielektromos oldhatósági modelleknek (CSMs) nevezzük.

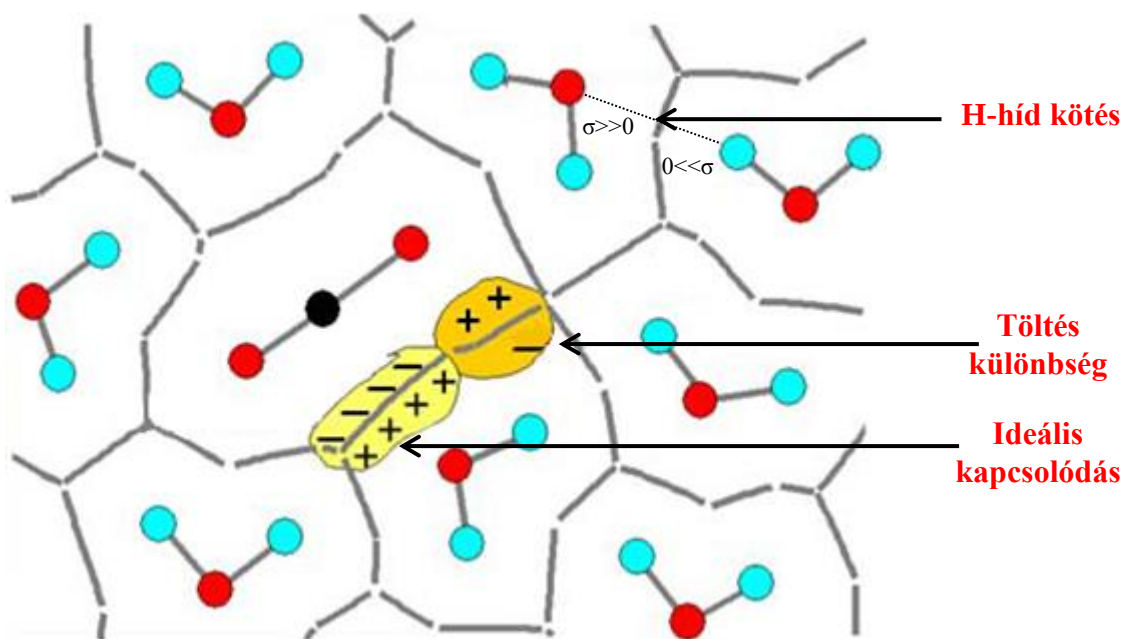
Az egyik legelterjedtebb CSMs modell az 1993-ben Klamt és Schüürmann [13] által kifejlesztett kvantumkémiai COSMO (Conductor-like Screening Model) elmélet. A COSMO elmélet szerint a molekulát egy - az alakját követő - szigetelő tömbfázissal (dielektrikum) határolt üregbe helyezzük. A legtöbb CSMs modell esetén csak a molekula töltéseloszlását (dipólus momentumait) tudjuk számítani. Sajnos a COSMO-nak is - mint az összes többi CSMs modellnek - vannak hátrányai, melyek korlátot szabnak alkalmazhatóságának. A módszer nem alkalmas elegyek és oldatok tulajdonságainak számítására. A molekula felületi töltéssűrűségnek meghatározásának alapjául klasszikus elektrosztatikus összefüggések (pontoszerű töltés) szolgáltak, ugyanakkor a molekula és az őt körülvevő dielektrikum között egy erősen nem-lineáris kapcsolat alakul ki. A CSMs modellek kizárólag egyszerűbb molekulák (pl. alkánok, víz) tulajdonságainak becslésére használhatók jól [14].

Szükség volt egy olyan megközelítésre, mely komplex molekulák és többkomponensű elegyek tulajdonságainak becslésére is használható. A probléma megoldására 1995-ben Klamt [15] kidolgozta a COSMO-RS elméletet (Conductor-like Screening Model for Real Solvents). A COSMO-RS egy fizikai modell, mely a különböző felületi töltéssűrűségű (σ -felület) molekulák kölcsönhatásait használja fel termodinamikai tulajdonságok meghatározására. Az *1. ábrán* egy molekula σ -felület látható. A σ -felület alkalmas a molekulák egyes részeinek kvalitatív elemzésére. Többféle dolog leolvasható: megfigyelhetők a molekula pozitív és negatív részei (piros és kék színek), a hidrogénhíd kötésre hajlamos részek (H-híd akceptorok a molekula erősen negatív részei, H-híd donorok erősen pozitív részei), illetve a molekula hidrofób és hidrofíl részei.



1. ábra: Egy példa molekula (Stavudine) σ -felülete

A COSMO-RS elmélet alapja, hogy a vizsgálat tárgyát képező molekulák felületi töltéssűrűség eloszlása ismert. Mindegyik molekula felülete tökéletesen illeszkedik egy másik molekula felületéhez, ennek következtében a COSMO-RS elmélet csak páros kölcsönhatásokat tud figyelembe venni. A teljes COSMO-RS érintkezési energia (E_{total}) három fő részből tevődik össze [14]. A különböző jelenségek bemutatása a 2. ábrán látható.



2. ábra: A teljes COSMO-RS érintkezési energiát (E_{total}) befolyásoló jelenségek bemutatása [14] (A van der Waals kölcsönhatások nincsenek ábrázolva)

Abban az esetben, ha az érintkező felületek töltésének (σ, σ') eredője nem nulla, akkor a lokális töltéskülönbség az 1. egyenlet segítségével írható le:

$$E_{\text{misfit}}(\sigma, \sigma') = a_{\text{eff}} \frac{\alpha'}{2} (\sigma + \sigma')^2 \quad (1)$$

ahol α_{eff} az érintkezési tényező a két felületegység között, illetve α' egy illesztendő paraméter. Az E_{misfit} lényegében az érintkező felületelemeken lévő eltérő töltésmennyiség miatt kialakuló elektrosztatikus kölcsönhatás miatt jön létre. Abban az esetben, ha az érintkező felületelemek felületi töltéssűrűsége között szélsőségesen nagy különbség van ($0 \ll \sigma$ felületelem lesz a H-híd akceptor, míg a $0 \gg \sigma$ felületelem a H-híd donor szerepet fogja felvenni), akkor a létrejövő H-híd kötés a 2. egyenlet szerint számolható:

$$E_{HB} = \alpha_{eff} C_{HB} \min \left[0; \min(0; \sigma_{donor} + \sigma_{HB}) \max(0; \sigma_{akceptor} + \sigma_{HB}) \right] \quad (2)$$

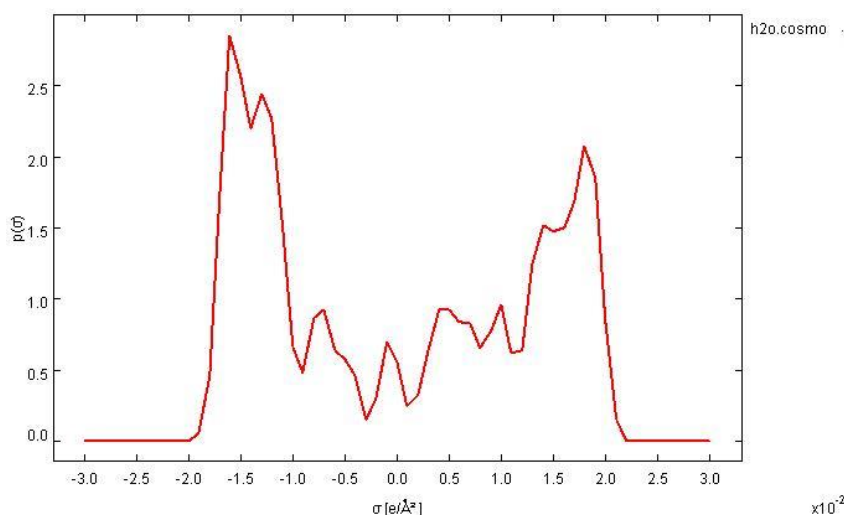
ahol C_{HB} és σ_{HB} illesztendő paraméterek. Az E_{total} értékét befolyásoló tényezők közül az utolsó, a molekulák között kialakuló van der Waals erő (3. egyenlet):

$$E_{vdW} = \alpha_{eff} (\tau_{vdW} + \tau_{vdW}') \quad (3)$$

ahol τ_{vdW} és τ_{vdW}' elemekre jellemző illesztendő paraméterek. Ezek alapján E_{total} értéke a 4. egyenlet felhasználásával számítható:

$$E_{total}(\sigma, \sigma') = E_{misfit}(\sigma, \sigma') + E_{HB} + E_{vdW} \quad (4)$$

Az eddig bemutatott egyenletek, mind a mikroszkópikus rendszerre (molekulák) voltak felírva. Ahhoz, hogy ez az információ a makroszkópikus szintre jusson, Klamt és munkatársai [16] a statisztikus termodinamikát hívták segítségül. Ehhez szükséges az ún. σ -profil fogalmának bevezetése. A σ -profil nem más, mint a kvantumkémiai COSMO számítás során meghatározott egységnyi felületre jutó töltés mennyiségének (E_{total}) és az adott töltéssel rendelkező egységnyi felületelemek számának összefüggése. A 3. ábrán a víz σ -profilja látható.



3. ábra: A víz σ -profilja

Megfigyelhető, hogy a görbe nagyon hasonló egy sűrűségfüggvényhez, azzal a kivétellel, hogy a görbe alatti terület nincs normálva. Ennek következménye, hogy kisebb molekulák esetén a σ -profil alatti terület kisebb, mint nagyobb molekulák esetén. A σ -profil matematikai jellemzésére az ún. σ -momentumok szolgálnak. Összesen 15 darab σ -momentumot definiáltak, melyek együttesen megfelelő módon írják le egy adott anyag σ -profilját [14]. A σ -momentumok három fő csoportba sorolhatóak: az első hét darab, a vegyület általános tulajdonságairól ad felvilágosítást (M_0 , M_1 , M_2 , M_3 , M_4 , M_5 , M_6), a következő 4 momentum, a H-híd kötés befogadására való hajlandóságot mutatja (HB_{acc1} , HB_{acc2} , HB_{acc3} , HB_{acc4}), az utolsó 4, pedig a H-híd kötés kezdeményezésére való hajlandóságról ad felvilágosítást (HB_{don1} , HB_{don2} , HB_{don3} , HB_{don4}). A momentumok között vannak, melyek valós fizikai jelentéssel bírnak, míg a többi momentumnak csupán matematikai jelentősége van [14]. Az M_0 momentum a molekula teljes felületének nagysága, az M_1 a molekula teljes COSMO polarizációs töltése, M_2 a molekula teljes COSMO elektrosztatikus kölcsönhatási energiája, M_3 a σ -profil forma mutatószáma, a HB_{acc3} , illetve HB_{don3} a molekula H-híd kötésre való hajlandóságát mutatja.

A σ -momentumok egyik lehetséges alkalmazása szerint feltételezzük azt, hogy a σ -momentumok és a becsülni kívánt fizikai-kémiai tulajdonság között lineáris kapcsolat van. Az 5. egyenlet egy általánosan felírható összefüggés, mellyel képesek vagyunk egy tetszőleges tulajdonság (P) becslésére σ -momentumok felhasználásával:

$$P = C_0 + C_1 \cdot M_0 + C_2 \cdot M_1 + C_3 \cdot M_2 + C_4 \cdot M_3 + C_5 \cdot M_4 + C_6 \cdot M_5 + C_7 \cdot M_6 + C_8 \cdot M_{HB_{acc1}} + C_9 \cdot M_{HB_{acc2}} + C_{10} \cdot M_{HB_{acc3}} + C_{11} \cdot M_{HB_{acc4}} + C_{12} \cdot M_{HB_{don1}} + C_{13} \cdot M_{HB_{don2}} + C_{14} \cdot M_{HB_{don3}} + C_{15} \cdot M_{HB_{don4}} \quad (5)$$

ahol $C_0, C_1, C_2 \dots C_{15}$ egy adott fizikai-kémiai tulajdonságra jellemző illesztendő paraméterek. A σ -momentumok korrelációs egyenletekben történő felhasználására számos szakirodalmi példa található. Mehler és munkatársai [17] a σ -momentumok lineáris kombinációját felhasználva becsülték meg különböző komponensek atkív szénen történő adszorpciós tulajdonságait. Winget és munkatársai [18] szintén adszorpciós tulajdonságok becslésére használták σ -momentumokat, ugyanakkor ebben az esetben adszorbensként különböző összetételű talajmintákat használtak.

A COSMO-RS számítás további lépéseiben feltételezzük, hogy az elegy minden egyes komponensének ismert a σ -profilja, vagyis kvantumkémiai COSMO számítást végeztünk az adott molekulára.

Többkomponensű elegyek esetén az egyik legfontosabb termodinamikai tulajdonság a komponensek kémiai potenciálja (μ_s). A COSMO-RS elmélet szerint az érintkező molekulák kémiai potenciálja egy adott felületegységen a 6. egyenlet egyenlet szerint számítható:

$$\mu_s(\sigma) = -kT \ln \int p_s(\sigma') \exp \left[-\frac{E_{total}(\sigma, \sigma') - \mu_s(\sigma')}{kT} \right] d\sigma' \quad (6)$$

ahol k a Boltzmann-állandó, T a hőmérséklet, $p_s(\sigma')$ a σ -profil. Látható, hogy μ_s mind a két oldalon szerepel, ezért az egyenlet megoldása csak egy iterációs ciklus segítségével lehetséges. A 7. egyenlet felhasználásával számítani tudjuk X komponens kémiai potenciálját az S rendszerben:

$$\mu_s^X = \int p^X(\sigma) \mu_s(\sigma) d\sigma + \mu_{C,S}^X \quad (7)$$

ahol $p^X(\sigma)$ az elegy σ -profilja, $\mu_{C,S}^X$ az adott rendszerre jellemző kombinációs tag, mely figyelembe veszi a molekulák méretének és alakjának eltérését. A kémiai potenciál ismeretében már számos termodinamikai tulajdonság számítható (fázisegyensúlyi jellemzők, sűrűség, gőznyomás stb.). Az egyik legfontosabb elegytulajdonság, a komponensek aktivitási tényezője a 8. egyenlet egyenlet segítségével számítható az egyes komponensek kémiai potenciáljának ismeretében:

$$\gamma_s^X = \exp \frac{\mu_s^X - \mu_x^X}{kT} \quad (8)$$

Látható, hogy a kvantumkémiai COSMO számítására épülő COSMO-RS elmélet segítségével számos termodinamikai tulajdonság becsülhető [19].

A COSMO-RS elmülethez kapcsolódó szakirodalmak nagy száma ellenére, az elmélet pontos alapjai nincsenek megfelelően közölve [20]. Ennek oka, hogy a kereskedelmi forgalomban megvásárolható szoftver (COSMOtherm) épül rá, melynek fejlesztői között Klamt és munkatársai is szerepelnek.

3. Az általam fejlesztett becslő módszer bemutatása

Célom egy olyan hőmérsékletfüggő gáz-folyadék határfelületi feszültség becslésére szolgáló becslő módszer (QSPR modell) kidolgozása volt, melynek bemenő paraméterei csupán az adott molekula szerkezetére jellemző paraméterek (σ -momentumok), és a hőmérséklet. Első lépésben a szakirodalom kritikus áttekintésével létrehoztam egy adatbázist, mely 178 vegyület gáz-folyadék határfelületi feszültség értékeit tartalmazza különböző hőmérsékleteken. A becslő módszer kidolgozásának következő lépésében kémiai rajzolóprogramok felhasználásával létrehoztam az adatbázisban szereplő molekulák szerkezetét. Ezt követően kvantumkémiai COSMO számítást végeztem, melynek eredményeként megkaptam a molekulák σ -momentumjait.

Első megközelítésként lineáris kapcsolatot tételeztem fel a σ -momentumok, illetve a hőmérséklet és a gáz-folyadék határfelületi feszültség között. Mivel igen sokváltozós szélsőérték keresési feladat jött létre, ezért többfajta minimukeresési eljárást is megvizsgáltam. A fejlesztendő becslő módszer pontosságát tovább növeltem egy nem-lineáris modellezési technika, a mesterséges neurális hálózat (ANN) alkalmazásával. A neurális hálózat optimalása során megvizsgáltam a neuronok átviteli függvényének, a neuronok és a rejtett rétegek számának hatását a kialakított becslő módszer pontosságára. Emelett kiválasztottam a neurális hálózat tanítására alkalmas algoritmust.

3.1 A bemeneti adatbázis létrehozása

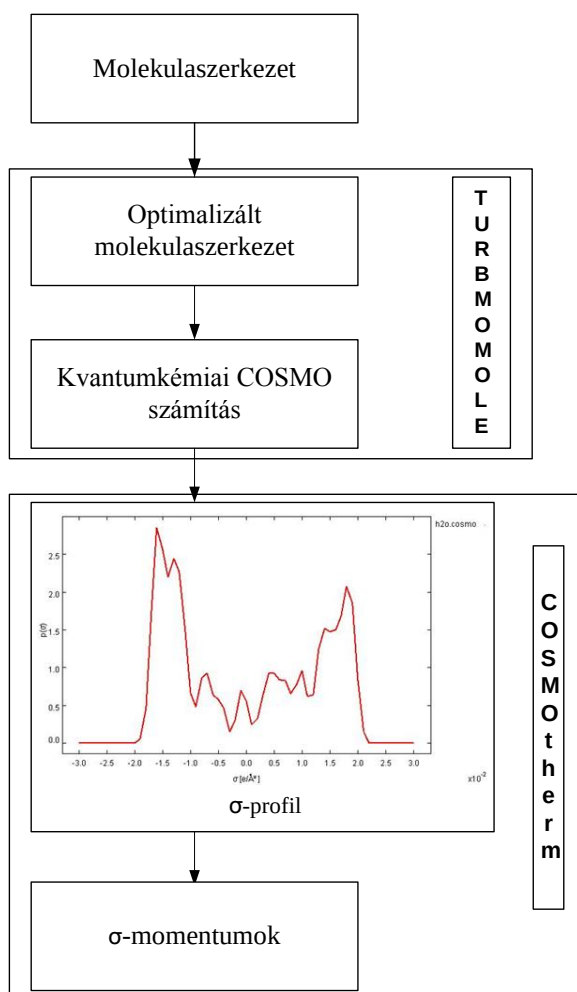
Számos szakirodalmi cikk tartalmaz kísérletileg meghatározott gőz-folyadék határfelületi feszültség értékeket. Ugyanakkor legtöbbször csak egyetlen vegyületre vagy kisebb vegyületcsaládra jellemző értékek kerülnek közlésre. Célom egy olyan irodalmi forrás kiválasztása volt, mely több vegyületcsaládra (akár elemekre is) tartalmaz kísérletileg meghatározott felületi feszültség értékeket. Jasper és munkatársai [21] által létrehozott adatbázis 1500 darab mérési pontot tartalmaz 178 vegyületre. Az

adastor diverzitását jól mutatja, hogy alkánokat, alkéneket, étereket, észtereket, alkoholokat, cikloalkánokat, aromásokat, aldehideket, ketonokat, aminokat, karbonsavakat, tiolokat, szerves halogénvegyületeket, egyéb szerves és ásványi savakat, illetve elemeket is egyaránt tartalmaz. Az adatsorban 0,07 és 45,08 [mNm⁻¹] gőzfolyadék határfelületi feszültség értékek között tartalmaz kísérletileg meghatározott adatpontokat. A szerző a mérési adatok pontosságáról nem ad egyértelmű felvilágosítást, viszont a becsült eltérés 0,1 és 0,2 [mNm⁻¹] értékek között van [21]. Ezen tulajdonságok alapján kijelenthető, hogy az adatsor jól reprezentálja a különböző szerves és szervetlen vegyületeket, így alkalmas egy QSPR alapú becslőmódszer bemeneti adatbázisának szerepére.

Az adatbázis digitális formában csak képként található meg. Ezért elsőként az adatokat át kellett konvertálnom olyan formátumban, amit már a Microsoft Excel segítségével is tudtam kezelni. A feladat megoldásához az OmniPage ingyenes felhőszolgáltatást alkalmaztam, mely alkalmas képeken karakterek felismerésére (OCR szoftver). Az így létrehozott adatbázist rendszereztem, majd elmentettem. Mivel igen sok mérési adatpontról van szó, ezért elengedhetetlen volt a folyamat automatizálása. A MATLAB egy teljesen felhasználóbarát magas szintű programozási nyelv, mely a mérnöki feladatok megoldása során rendkívül jól alkalmazható. Az adatbázis kialakítását, illetve az egyes feladatrészek megoldását MATLAB környezetben végeztem.

Mivel a fejlesztett becslőmódszer bemeneti adatai az egyes vegyületek σ -momentumai, ezért ezeket a megfelelő eljárás segítségével elő kellett állítanom. A vegyületek 2D-s szerkezetét irodalmi források felhasználásával állapítottam meg. Ezt követően az egyes molekulák ún. Z-mátrixát az Avogadro 1.0.1-es verziójú kémiai molekulaépítő program segítségével hoztam létre. A kvantumkémiai jól ismert Z-mátrix lényegében a molekulát alkotó atomok 3D-s Descartes-féle koordináta-rendszerbeli koordinátáit tartalmazza. Az így létrejött fájlokat a megfelelő fájlformátummá konvertáltam az OpenBabel 2.3.1-es verzióját használva. Az OpenBabel egy nyílt forráskódú konvertáló program, mely több mint 110 féle kémiai fájlformátumot ismer. Ezután a molekulák szerkezetének optimalizálásához és a kvantumkémiai COSMO számításhoz a 2013-ban megjelent TURBOMOLE 6.5 [22] szoftvert használtam. A folyamat eredményeként az ún. cosmo-fájlokat kaptam, melyekből az egyes vegyületek σ -momentumait a COSMO-RS elméletre épülő COSMOTHERM programcsomag segítségével számítottam ki. A

számítók során a C21_0110 verziójú programot használtam a BP_TZVP_C21_1010 jelű parametrizálással [23]. A számítás menete a 4. ábrán látható:



4. ábra: A σ -momentumok előállításának menete

3.2 Többváltozós lineáris regresszió

A többváltozós lineáris regressziót, mint QSPR modellt a szakirodalomban több szerző is javasol [24-27]. Ezért a becslőmódszer kidolgozásának következő lépéseként egy olyan összefüggést kellett találnom, mely pontosan képes leírni σ -momentumok és a hőmérsékletfüggő felületi feszültség közötti kapcsolatot. A többváltozós lineáris regresszió létrehozása során célom az volt, hogy megvizsgáljam, a σ -momentumok és a hőmérsékletfüggő gáz-folyadék határfelület közötti kapcsolat leírható-e egy többváltozós lineáris regresszió segítségével. Elsőként a többváltozós lineáris

(multilineáris) regressziót alkalmaztam. A felállított modell egyenlet (9. egyenlet) a következő:

$$\gamma = C_0 + C_1 \cdot M_0 + C_2 \cdot M_1 + C_3 \cdot M_2 + C_4 \cdot M_3 + C_5 \cdot M_4 + C_6 \cdot M_5 + C_7 \cdot M_6 + C_8 \cdot M_{Hbacc1} + C_9 \cdot M_{Hbacc2} + C_{10} \cdot M_{Hbacc3} + C_{11} \cdot M_{Hbacc4} + C_{12} \cdot M_{Hbdon1} + C_{13} \cdot M_{Hbdon2} + C_{14} \cdot M_{Hbdon3} + C_{15} \cdot M_{Hbdon4} + C_{16} \cdot T \quad (9)$$

ahol $C_0, C_1, C_2 \dots C_{16}$ a hőmérséklet függő gáz-folyadék határfelületi feszültség és a σ -momentumok kapcsolatának leírására szolgáló illesztendő paraméterek. Látható, hogy a fenti egyenlet nagyon hasonló az 5. egyenlethez, viszont ebben az esetben a σ -momentumok mellett a hőmérséklet is szerepel, mint bemeneti változó. Ebben az esetben mind a 15 σ -momentumot felhasználtuk a gáz-folyadék határfelületi feszültség leírására. Járvás és munkatársai [28] bemutatták, hogy 5 meghatározott σ -momentum közel ugyanannyi információt hordoz, mint a 15 σ -momentum. Ezt az 5 momentumot ($M_0, M_2, M_3, M_{Hbacc3}, M_{Hbdon3}$) ún. Klamt szettnek nevezzük (a COSMO-RS elmélet kidolgozója után). Ebben az esetben a 10. egyenlet segítségével lehet leírni a hőmérsékletfüggő gáz-folyadék határfelületi feszültség és a σ -momentumok közötti kapcsolatot :

$$\gamma = C_0 + C_1 \cdot M_0 + C_2 \cdot M_2 + C_3 \cdot M_3 + C_4 \cdot M_{Hbacc3} + C_5 \cdot M_{Hbdon3} + C_6 \cdot T \quad (10)$$

Látható, hogy a 10. egyenlet használatával az illesztendő paraméterek száma a harmadára csökkent. A minimumkeresési eljárást MATLAB környezetben hajtottam végre. A minimalizálandó célfüggvényként a kísérletileg meghatározott és számított gáz-folyadék határfelületi feszültség négyzetes eltérését vettem. A négyzetes eltérés előnye, hogy a kis eltéréseket kevésbé veszi figyelembe, ugyanis 1-nél kisebb szám négyzete még kisebb lesz. Nagyobb eltérések esetén viszont felnagyítja a hibát, gyakorlatilag egy büntető függvényként funkcionál. Mivel sokváltozós minimumkeresési eljárásról van szó, ezért fontos az optimálisan kiválasztott minimumkeresési eljárás. Elsőként Nelder-Mead-féle módosított szimplex módszerrel [29] alapuló determinisztikus minimumkeresést alkalmaztam. A determinisztikus keresési eljárások problémája, hogy az esetleges lokális minimumok jelentősen megzavarhatják a globális minimum megtalálását. Gyakorlatilag ez azt jelenti, hogy a minimumkeresési eljárást különböző kezdőpontokból indítva, eltérő végeredményt kapunk. Léteznek ún. sztochasztikus minimumkeresési eljárások, melyek célja, hogy véletlen kiindulási ponttól kezdve feltérképezzék az egész keresési teret. Jellemzőjük, hogy nem igényelnek kezdőpontot, és eredményként - optimális működés esetén - a

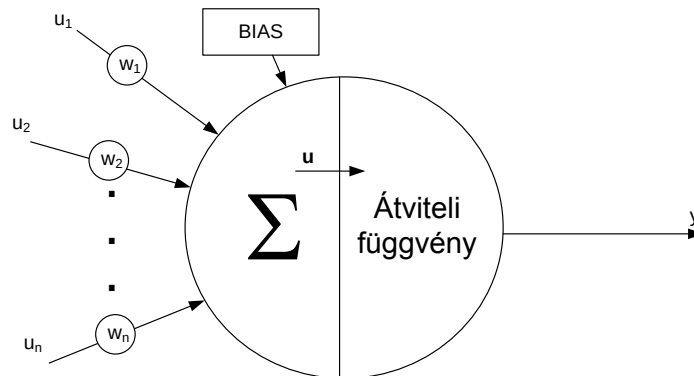
globális minimumot szolgáltatják. A sztochasztikus minimumkeresési eljárások csoportjába tartoznak az ún. evolúciós keresési stratégiák, melyek közül a CMA-ES-t (Covariance Matrix Adaptation Evolution Strategy) [30] alkalmaztam. Az evolúciós stratégiák jellemzője, hogy minden egyes keresési lépésben közre játszanak „öröklődési”, illetve „mutálódási” folyamatok a következő lépésbeli kiindulási pontok meghatározásakor. Mivel az eddig bemutatott két minimumkeresési eljárás esetén nem sikerült megfelelő pontosságú összefüggést létrehoznom, ezért egy újabb sztochasztikus minimumkeresési eljárást is bevontam a vizsgálatban. Az ún. részecske-raj optimalizáció (PSO - Particle Swarm Optimization) természeti megfigyeléseken alapul [31]. A keresési térben elindított „egyedek” célja a teljes tér végigtapogatása a lehető leghatékonyabb módon. Az egyes egyedek mozgása egy megadott mozgásegyenlet segítségével írható le. A mozgásegyenletben szerepelnek az egyedek közti kölcsönhatások (a más részecskék által érintett minimumtól való távolság), az adott egyed tehetetlensége, illetve az egyed egyéni hatása is (saját maga által érintett minimumtól való távolság).

Sajnos a felhasznált minimumkeresési eljárásokkal nem sikerült megfelelő illeszkedését elérnem, vagyis feltételezhetően nem-lineáris összefüggés van a σ -momentumok és a hőmérsékletfüggő gáz-folyadék felületi feszültség között. Katritzky és munkatársai [32] kimutatták, hogy a QSPR modellek nagyon ritkán lineáris viselkedésűek, a tapasztalt jelenségek megfelelő pontosságú leírására valamilyen nem-lineáris leírási technikát kell választanunk. Ugyanakkor a későbbi felhasználás szempontjából, egy lineáris összefüggés használata igen egyszerű és jól algoritmizálható.

3.3 Mesterséges neurális hálózatok

A mesterséges neurális hálózat (ANN), mint nem-lineáris modellezési módszer, a QSPR modellekhez kapcsolódó szakirodalomban igen sokszor előfordul [33-38]. A mesterséges neurális hálózatok a fekete-doboz modellezési technikák (a modellben alkalmazott paramétereknek nincs valós fizikai jelentése) közé tartoznak, melyek olyan esetekben alkalmazhatók jól, amikor nem ismerjük a leírandó összefüggés pontos felépítését. A neurális hálózat fejlődése az 1950-es években indult meg. Kezdetben pusztán biológiai szempontból vizsgálták az agyban lévő neuronok működését. Megállapították, hogy a neuron működésének központi egysége a neuron sejtmagja, melyhez a szinapszisokon keresztül az axonok jutattják el az elektromos impulzust.

Rájöttek arra is, hogy az agyban található neuronok két állapotúak (vagy adnak jelet vagy nem). A mesterséges neurális hálózatok jelentősen hasonlítanak a biológiai neurális hálózatokhoz, viszont összetettségük sokkal kisebb. Még egy biológiai neurális hálózat több milliárd neuronból áll, addig mesterséges társa csupán néhány neuront tartalmaz [39]. A mesterséges neurális hálózat működésének megértéséhez elsőként meg kell vizsgálnunk egy neuron működését (5. ábra).

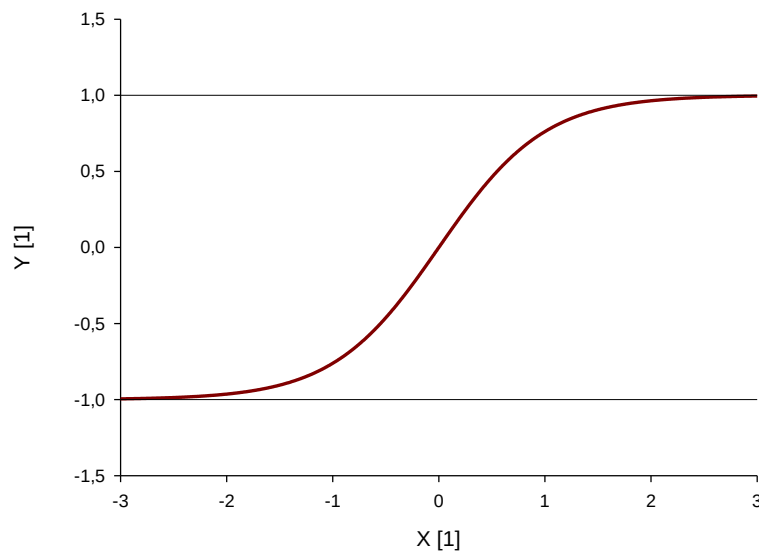


5. ábra: Egy mesterséges neuron működése (n db bemenet, 1 db kimenet)

Az 5. ábrán látható, hogy a neuron két fő részből áll. A bemeneti oldalon helyezkedik el a bemenetek összegzésére szolgáló rész, míg a kimeneti oldalon egy átviteli függvény található (a szakirodalomban előfordul aktivációs függvény néven is). A neuron összesített bemenetét (u) a különböző bemeneti változók (u_i) és a hozzájuk tartozó súlyok (w_i) szorzata alapján kaphatjuk meg (11. egyenlet):

$$u = \sum_{i=1}^n u_i w_i \quad (11)$$

Legtöbbször a neuron bemenetei között szerepel egy konstans tag (BIAS - „zavarás”), melyhez külön súly nem tartozik. Az így kapott összegzett bemenet az átviteli függvénybe kerül. Az átviteli függvények legtöbbször korlátos függvények, de használhatunk lineáris átviteli függvényt is. Legtöbbször valamilyen korlátos hiperbólikus függvényt alkalmazunk átviteli függvényként. Az egyik legjobban alkalmazható átviteli függvény a tangens hiperbolikus alakja a 6. ábrán látható.

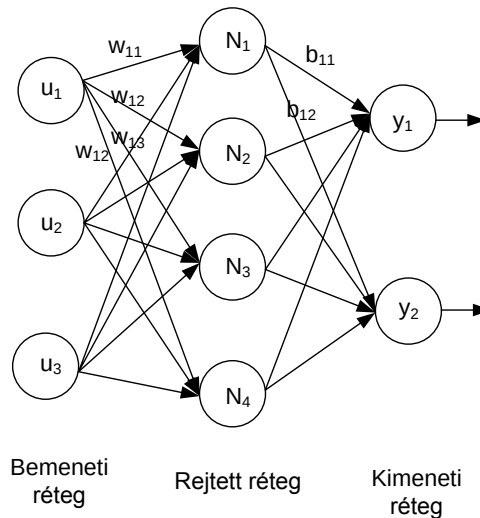


6. ábra: Tangens hiperbólikusz alakú átviteli függvény

Ezen kívül használhatunk lépcsőfüggvényt, telítési lineáris függvényt vagy logisztikus (logaritmus hiperólikus függvénye) függvényt is. A neuron kimenet a 12. egyenlet segítségével számítható:

$$y = f(u + BIAS) \quad (12)$$

ahol f a felhasznált átviteli függvény. Mint látható, önmagában egy darab neuron nem képes a bemeneti, illetve kimeneti változók közötti kapcsolat megfelelő pontosságú leírására, ezért a neuronokat össze kell kapcsolnunk. A neuronok összekapcsolódásából származó rendszert nevezzük mesterséges neurális hálózatnak. Egy 3 bemenetű, illetve 2 kimenetű mesterséges hálózat felépítése a 7. ábrán látható:



7. ábra: Mesterséges neurális hálózat felépítése (példa). Egyszerűsítés képpen csak az első neuronhoz tartozó súlyok lettek feltüntetve

A 7. ábrán megfigyelhető egy tipikus neurális hálózat topológiája. Munkám során az előrecsatolt mesterséges hálózatokkal foglalkoztam. Az előrecsatolás ebben az esetben azt jelenti, hogy adott rétegekből jövő információ mindig az azt követő rétegbe kerül, vagyis az azt megelőző rétegbe információ nem juthat vissza. A bemeneti rétegben elhelyezkedő neuronok lényegében csak a bemeneti változót szolgáltatják a hálózat számára, átviteli függvényt, összegző modult, illetve BIAS-t ezek a neuronok nem tartalmaznak. A rejtett rétegben, illetve a kimeneti rétegben elhelyezkedő neuronok felépítése, gyakorlatilag megegyezik az 5. ábrán láthatóéval. Mindkét rétegben lévő neuronok tartalmaznak átviteli függvényt, összegző modult és BIAS-t is. Ugyanakkor, míg a rejtett rétegben lévő neuronok tetszőleges átviteli függvényt tartalmazhatnak, addig a kimeneti rétegbeli neuronokba csak lineáris átviteli függvényt építhetünk. Látható, hogy a neuronok közötti kapcsolat kizárólag vertikális, vagyis az egy rétegben elhelyezkedő neuronok között nincsen kapcsolat. Egy általános mesterséges neurális hálózat egy-egy darab bemeneti, illetve kimeneti réteget és tetszőleges számú rejtett réteget tartalmazhat. A bemeneti rétegben lévő neuronok számát a hálózat bemeneteinek száma, még a kimeneti rétegben elhelyezkedő neuron számot a kimenetek száma határozza meg. A Cybenko-tétel [40] értelmében, minden nem-lineáris probléma leírható maximum három rejtett réteg használatával. A neurális hálózatok legnagyobb előnye a neuronok különböző jellemzőinek (neuronok és rejtett rétegek száma, neuronok átviteli függvénye, tanítási algoritmus stb.) variálhatóságában rejlik.

A mesterséges neurális hálózat struktúrájának létrehozása után, a hálózat tanítási folyamata következik. A tanítás során az előzőleg létrehozott, bemeneti, illetve a kísérletileg meghatározott kimeneti adatokat tartalmazó adatbázis egy részének felhasználásával történik. A tanítási folyamat lényege az egyes kapcsolatokhoz tartozó súlyok (w_{ij}, b_{ij}) meghatározása. Lényegében egy rendszer identifikációról beszélhetünk, aminek célja a kísérletileg meghatározott és a mesterséges neurális hálózat által számított kimeneti értékek négyzetes eltérésének összegének minimalizálása, a súlyok változtatása mellett. A BIAS értéke a tanítás folyamata során véletlenszerűen változik, ezért egy ilyen típusú hálózat tanítása ciklikus folyamat. Fontos, hogy a tanítási adatbázis hasonlóan heterogén összetételű legyen, mint a teljes adatbázis. A minimumkeresési eljárás bonyolultságát jól mutatja, hogy a 7. ábrán látható hálózat esetén a keresési változók száma 20 (12 darab súly a bemeneti oldal és a rejtett réteg, míg 8 darab súly a rejtett réteg és a kimeneti réteg között). A mesterséges neurális hálózat tanítását a kimeneti rétegtől kell kezdenünk (backpropagation), majd így haladunk a bemeneti réteg irányában. Többféle minimumkeresési eljárást megvizsgáltam. Módosított gradiens módszerek közül a Polak-Ribière (traincgp) [41], a Fletcher-Reevesv (traincgf) [42], illetve a Powell-Beale (traincgb) [43] algoritmusokat. Ezek a módszerek csupán kis mértékben térnek el egymástól. A részletes bemutatástól eltekintve elmondható, hogy mindhárom módszer kis mértékben különböző képen határozza meg a következő keresési pont helyét a legnagyobb meredekség (gradiens) irányában. Ezen kívül megvizsgáltam a legkisebb négyzetek módszerére épülő Levenberg-Marquardt (trainlm) tanítási algoritmust [44]. A vizsgálatok célja az optimális tanítási algoritmus kiválasztása volt. A tanítás végén egy olyan mesterséges neurális hálózattal rendelkezünk, mely tartalmaz minden - a leíráshoz szükséges – súlyt (w_{ij}, b_{ij}), illetve BIAS értéket.

A mesterséges neurális hálózat tanítása után validálnunk kell az hálózatot. Ez azt jelenti, hogy a bemeneti, illetve kísérletileg meghatározott adatbázis egy kis részét felhasználva megvizsgáljuk a hálózat által számított és kísérletileg meghatározott kimeneti változók eltérését. A validálási lépés gyakorlatilag a neurális hálózat tanításához tartozik. Célja, hogy elkerüljük az ún. túltanítás jelenségét, mely akkor lép fel, ha neurális hálózat túlságosan „jó” leírja azt az adatsort, amivel tanítottunk, ugyanakkor más adatsorok esetén nem ad jó eredményt. A validálást sok esetben ki lehet hagyni a neurális hálózat

létrehozásának lépései közül, viszont ahol viszonylag nagy adatbázis áll rendelkezésünkre, ott szükséges a belső validálási lépés beiktatása.

A neurális hálózat létrehozásának utolsó művelete a hálózat tesztelése, mely során egy - a tanítási és validálási csoporttól - független adatbázis esetén hasonlítjuk össze a kísérletileg meghatározott kimenetek és a hálózat által becsült kimenetek eltérését. A tesztelés eredménye alapján kapunk felvilágosítást arról, hogy a hálózat hogyan működik olyan adatsorok esetén, melyekkel a tanítás és a validálás során nem találkozott.

A rendelkezésemre álló 1500 adatpontot tartalmazó adatbázis 70%-át (1050 db adatpont) tanításra, illetve 15-15% át (225 db adatpont) belső validálásra, illetve tesztelésre használtam. Fontos, hogy a különböző csoportokban lévő adatsorok hasonló heterogenitással bírjanak, ezért az adatbázisban szereplő vegyületeket kémiai jelleg szerint szétosztottam a különböző csoportok között. Célom az volt, hogy lehetőleg mindegyik vegyületcsaládból kerüljön valamennyi csoportba, különös figyelemmel a tanítási és teszt csoportokra. Az első 1050 adatpontot a tanítási csoportba, a következő 225-225 darabot pedig a validálási, illetve tesztelés csoportokba helyeztem. A mesterséges neurális hálózatot az egyes csoportokhoz tartozó négyzetes korrelációs koefficiens (R^2) segítségével minősítettem. Abban az esetben, ha mindhárom csoporthoz tartozó R^2 értéke megfelelően nagy, akkor a hálózat rendelkezik a hőmérsékletfüggő gáz-folyadék határfelület becsléséhez szükséges predikációs képességgel.

A hőmérsékletfüggő gáz-folyadék határfelület becslésére szolgáló módszer létrehozása mellett, munkám célja volt annak a molekuláris jellemzőnek a kiválasztása, amely legnagyobb mértékben határozza meg a gáz-folyadék határfelületet feszültséget. Lineáris rendszerek esetén a főkomponens analízis (PCA) használható annak a bemeneti változónak a kiválasztására, mely legnagyobb hatással van a kimenetek értékére. Nem-lineáris rendszerek esetén a NLPCA (nem-lineáris főkomponens analízis) alkalmazható, viszont ebben az esetben ismernünk kell az összefüggést, a bemeneti, illetve a kimeneti változók között. Vagyis rendelkezünk kell egy egyenlettel, mely összekapcsolja a bemeneti és kimeneti változókat. Sajnos ebben az esetben nem rendelkeztem ilyen, egyértelmű összefüggéssel, ugyanakkor rendelkezésemre állt az optimalizált mesterséges neurális hálózat, melynek segítségével meg tudtam határozni az egyes bemeneti változók „fontosságát”. Az ún. súlyok módszerének [45] segítségével meg

tudtam állapítani, mely bemeneti változó bír a legnagyobb „súllyal” a gáz-folyadék határfelületi feszültség meghatározásában. Ehhez a rejtett rétegbeli neuronokhoz tartozó súlyokat összeadtam bemenetenként, majd megszoroztam a kimeneti réteghez tartozó súlyokkal. Így megkaptam az egyes bemenetekhez tartozó teljes súlyt a hálózatban. Az ún. perturbációs módszer [46] lényege, hogy az optimalizált hálózat bemenetét egyenként százalékosan megnöveltem. A különböző bemenetek változtatása mellett figyeltem a neurális hálózat kimenetét (gáz-folyadék felületi feszültség). Amelyik bemenet változtatása esetén legnagyobb az eltérés az optimalizált hálózat által számított kimenethez képest, az a bemeneti változó befolyásolja legnagyobb mértékben a hálózat kimenetét. Az ún. COM módszer [47] alapján is elvégeztem a bemenetek fontosságának rangsorolását. A módszer alapja, hogy egy adott bemeneti változó kihagyása mellett végezzük el a mesterséges neurális hálózat létrehozását. Tehát ebben az esetben 5 bemeneti változó (4 momentum és a hőmérséklet) áll rendelkezésre. Amelyik változó esetén legnagyobb mértékben tér el az új hálózat R^2 értéke az eredeti optimalizált hálózatétól, az a változó befolyásolja legnagyobb mértékben a hőmérsékletfüggő gáz-folyadék felületi feszültséget. Az R^2 értékek mellett a vizsgálatok során az ún. MAE értékét is figyeltem, mivel sok esetben az R^2 értéke nem ad kellő felvilágosítást. A MAE-t a következő képlet szerint számítottam (11. egyenlet):

$$MAE = \frac{\sum_{i=1}^N (\gamma_{\text{becsült}}^i - \gamma_{\text{kísérleti}}^i)^2}{N} \quad (11)$$

ahol N a mérési pontok száma.

A mesterséges neurális hálózatok létrehozását MATLAB 2010b szoftverkörnyezetben végeztem saját fejlesztésű algoritmus segítségével. A számításokat egy HP Proliant G6 szerveren végeztem. Az adatok kimentéséhez a Microsoft Office programcsomagot vettem igénybe.

4. Eredmények

Az eredmények bemutatása során ismertetésre kerülnek a többváltozós lineáris regresszió eredményei. Ezt követően a mesterséges neurális hálózat optimalizálásának lépései következnek, majd bemutatom az optimalizált mesterséges neurális hálózattal végzett becslés eredményét. Végül az optimalizált neurális hálózat segítségével

bemutatok egy lehetséges molekuláris szintű magyarázatot a gáz-folyadék határfelületi feszültség jelenségére.

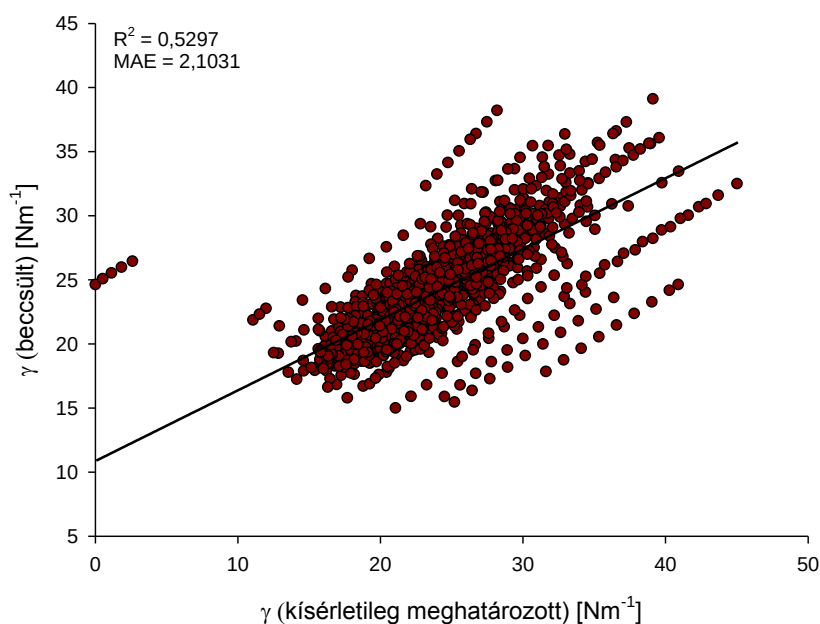
4.1 Többváltozós lineáris regresszió

A többváltozós lineáris regresszió során lineáris összefüggést tételeztem fel a σ -momentum és a hőmérsékletfüggő gáz-folyadék határfelületi feszültség között. Elsőként a teljes σ -momentum szettet (15 darab) használtam az összefüggés létrehozásához (9. egyenlet). A különböző minimumkeresési eljárások alkalmazásával elért eredményeket, 15 momentum felhasználása mellett, a 1. táblázat tartalmazza:

1. táblázat: Többváltozós lineáris regresszió eredménye 15 momentum esetén

	MAE	R ²
Módosított szimplex módszer	2,1791	0,5083
CMA-ES	2,1168	0,5290
PSO	2,1031	0,5297

Az 1. táblázatban látható, hogy a legkisebb MAE, illetve legnagyobb R² értékét a részecske-raj optimalizációs eljárás esetén tapasztaltam. Ez azt mutatja, hogy igen jól használható eljárásról van szó, mely sok esetben a mérnöki feladatmegoldás szerves részét képezi. Fontos kiemelni, hogy a PSO esetén kaptuk a legkisebb MAE értéket, mely nem arányosan csökkent az R²-ek értékének növekedésével. Ez azt mutatja, hogy az R² és MAE együttesen adnak felvilágosítást a korreláció erősségére vonatkozóan. A 8. ábra a legjobbnak talált korreláció eredményét mutatja 15 momentum esetén:



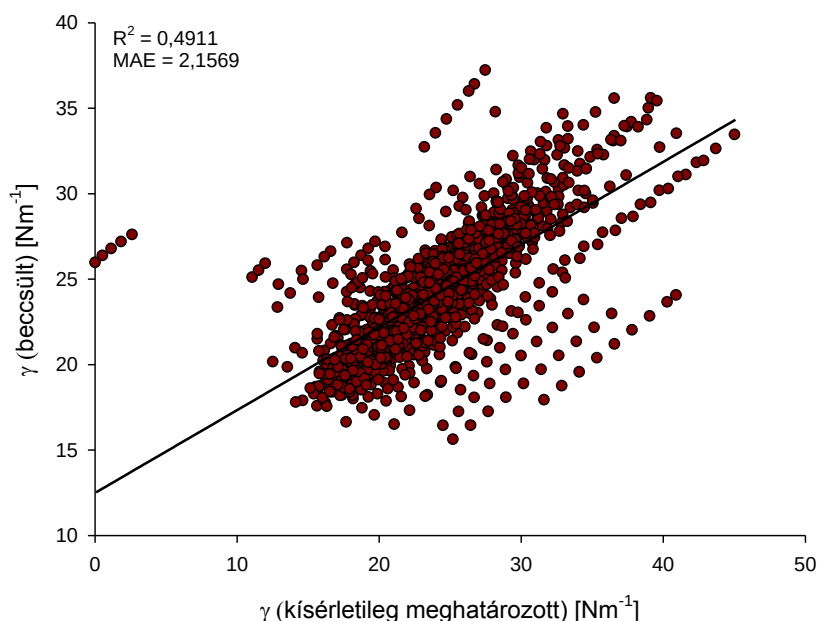
8. ábra: A többváltozós lineáris regresszió eredménye 15 momentum esetén (PSO)

A 8. ábrán látható, hogy a többváltozós lineáris regresszió 15 momentum felhasználásával nem képes jól leírni a hőmérsékletfüggő gáz-folyadék határfelületi feszültséget. A vegyületek egy részénél ugyan jó közelítést adhat ez a módszer, ugyanakkor legtöbbször nem alkalmas a gáz-folyadék felületi feszültség becslésére. Az ábrán megfigyelhető trendek, az adott vegyület különböző hőmérsékleten mért gáz-folyadék határfelületi feszültségéből adódnak. Következő lépésben megvizsgáltam, hogy az ún. Klamt szettet felhasználva, milyen eredményt kapok. Ebben az esetben a 10. egyenlet felhasználásával végeztem el a többváltozós lineáris regressziót. A 2. táblázat a különböző minimumkeresési eljárások alkalmazásával kapott eredményeket mutatja a Klamt szett esetén:

2. táblázat: Többváltozós lineáris regresszió eredménye a Klamt szett (5 darab momentum) esetén

	MAE	R ²
Módított szimplex módszer	2,5523	0,4662
CMA-ES	2,1522	0,4091
PSO	2,1569	0,4911

A 2. táblázatban látható, hogy ebben az esetben is a PSO eljárás produkálta a legmagasabb R² értékét, viszont a legalacsonyabb MAE értéket a CMA-ES eljárásnál tapasztaltam. Fontos kiemelni, hogy a R² és MAE értékek az 1. táblázathoz képest csak kis mértékben csökkentek. Ez alapján kijelentő, hogy a Klamt szett a gáz-folyadék határfelületi feszültség becslése során közel azonos kémiai információt kódol, mint a 15 momentum. A 9. ábrán a regresszió eredménye látható a Klamt szett esetében:



9. ábra: A többváltozós lineáris regresszió eredménye a Klamt szett esetén (PSO)

A 9. ábrán látható, hogy ebben az esetben sem rendelkezik a modell megfelelő predikciós képességgel. Következétként levonható, hogy a hőmérsékletfüggő gáz-folyadék határfelület becslésére szolgáló modell létrehozásához nem kell mind a 15 momentumot bevonni, ugyanis a Klamt szett (M_0 , M_2 , M_3 , M_{Hbacc3} , M_{Hbdon3}) hordozza a szükséges információkat.

A többváltozós lineáris regresszió eredményei alapján kijelentő, hogy a gáz-folyadék határfelületi feszültség és a σ -momentumok között nem lineáris az összefüggés, ezért a továbbiakban valamilyen nem-lineáris regressziós technikát kell alkalmazni. A feladat elvégzésére a mesterséges neurális hálózatok bizonyultak a legkézenfekvőbb megoldásnak.

4.2 Mesterséges neurális hálózatok

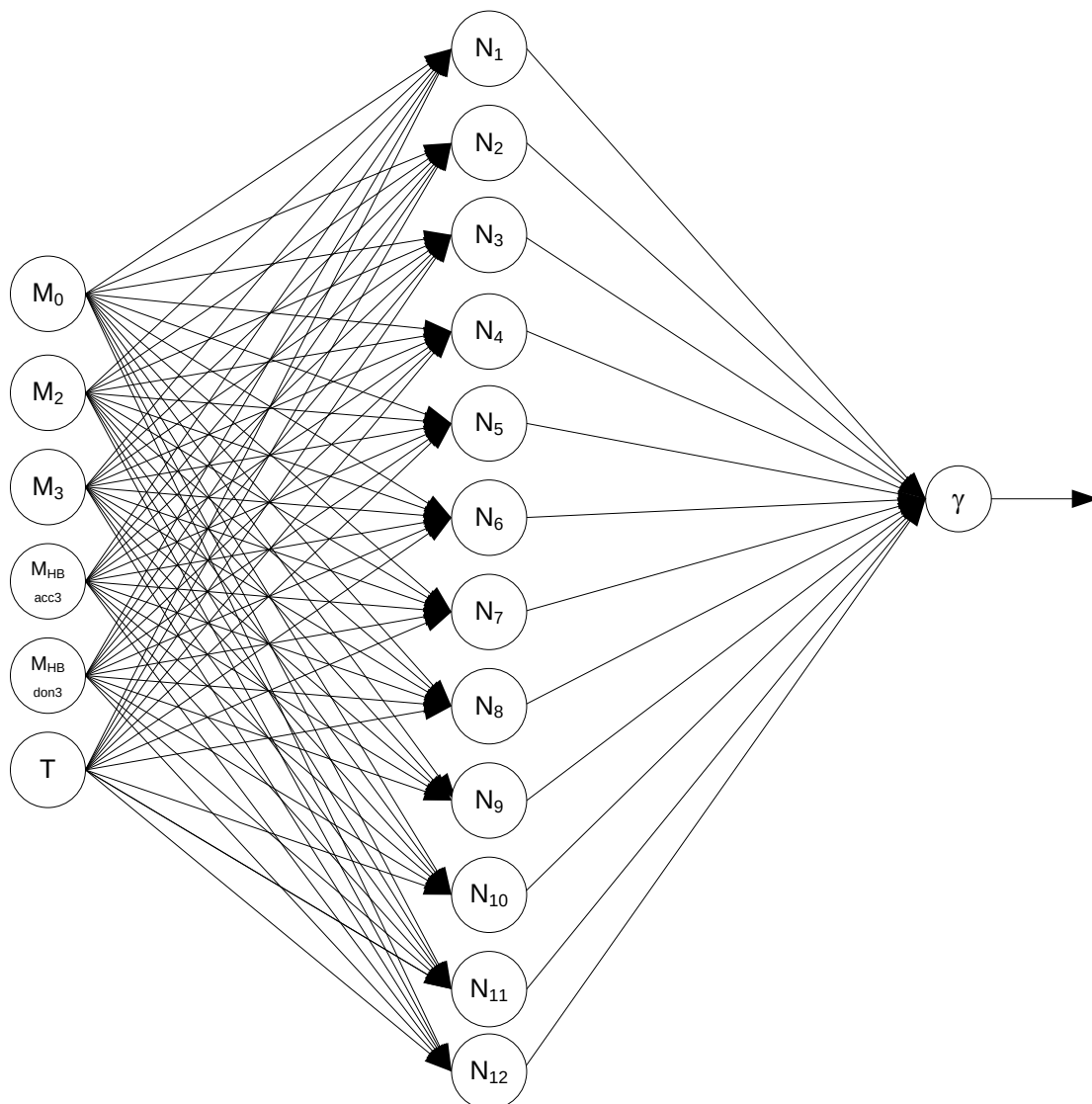
A mesterséges neurális hálózat létrehozásakor célom volt, hogy a lehető legjobb predikciós képességgel rendelkező gáz-folyadék határfelületi feszültség becslésére szolgáló nem-lineáris modellt hozzam létre. A többváltozós lineáris regresszió során kimutattam, hogy elegendő a Klamt szettet és a hőmérsékletet figyelembe vennem, mint bemeneteket, a becslőmódszer kidolgozása során. Egy mesterséges neurális hálózat számos különböző tulajdonsággal jellemezhető: topológia, átviteli függvény, illetve a tanításhoz használt algoritmus. Elsőként az egy rejtett rétegben elhelyezkedő neuronok optimális számát határoztam meg. A vizsgálat eredményét a következő táblázat tartalmazza:

3. táblázat: Egy rejtett rétegben szereplő neuronok számának optimalizálása

	R^2 (tanítás)	R^2 (validálás)	R^2 (teszt)
14 neuron	0,9738	0,9559	0,9679
12 neuron	0,9665	0,9326	0,9607
10 neuron	0,9429	0,9294	0,9431
8 neuron	0,9292	0,9035	0,9384
5 neuron	0,8695	0,8326	0,8717
3 neuron	0,7795	0,7754	0,7658

A 3. táblázatban látható, hogy a neuronok számának növekedésével a mesterséges neurális hálózathoz tartozó R^2 értékek is növekednek. Fontos megállapítani, hogy már 3 neuron esetén jobb eredményt érünk el, mint többváltozós lineáris regresszió segítségével. A neuronok száma nem lehet túlságosan nagy, ugyanis ebben az esetben a hálózat létrehozása (tanítás, validálás és tesztelés) lényegesen tovább tart. Ezen okok miatt, egy rejtett rétegben 12 darab neuront helyeztem el. Mivel a számítások eredménye már egy rejtett réteg esetén is megfelelően nagy, ezért a rejtett rétegek számát egynek választottam. Több rejtett réteg esetén a hálózat létrehozásához

szükséges idő nagyságrendileg nagyobb. A kialakított neurális hálózat végleges topológiáját a következő ábra mutatja:



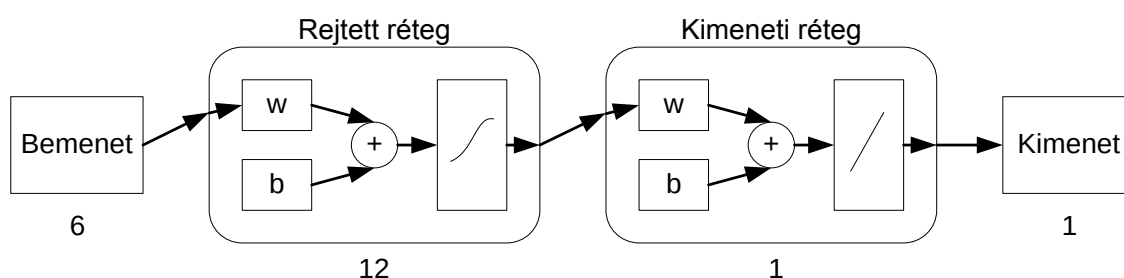
10. ábra: A hőmérsékletfüggő gáz-folyadék határfelületi feszültség becslésére szolgáló neurális hálózat topológiája

Következő lépésben a neuronok átviteli függvényét határoztam meg. A vizsgálat eredményét a következő táblázat (4. táblázat) tartalmazza:

4. táblázat: A rejtett rétegbeli neuronokban lévő átviteli függvény kiválasztása

	R^2 (tanítás)	R^2 (validálás)	R^2 (teszt)
tangens hiperbólikusz	0,9649	0,9471	0,9628
lineáris	0,5048	0,4977	0,4815
logisztikus	0,9635	0,9361	0,9584

A 4. táblázatban látható, hogy legjobb eredményt tangens hiperbolikus átviteli függvény esetén kaptam. Érdekes megjegyezni, hogy lineáris átviteli függvény esetén hasonló eredményt kaptam, mint a többváltozós lineáris regresszió során. Ennek oka, hogy a neurális hálózatban lévő súlyokat összevonva, megkaphatjuk a többváltozós lineáris egyenletben szereplő konstans tagokat. A 11. ábrán látható az optimalizált neurális hálózat egyszerűsített ábrája:.



11. ábra: A gáz-folyadék határfelületi feszültség becslésére szolgáló mesterséges neurális hálózat egyszerűsített ábrája (MATLAB ábra)

Következőként kiválasztottam a megfelelő tanítási algoritmust. A vizsgálat eredményét a következő táblázat (5. táblázat) tartalmazza:

5. táblázat: A rejtett rétegbeli neuronokban lévő átviteli függvény kiválasztása

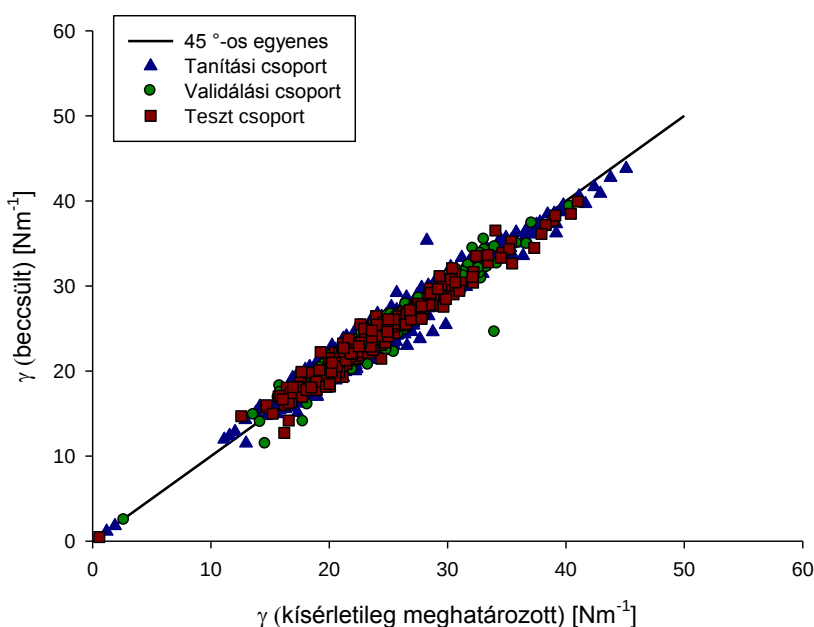
	R^2 (tanítás)	R^2 (validálás)	R^2 (teszt)
traincgp	0,8173	0,8173	0,7991
traincgf	0,8074	0,8282	0,7947
traincgb	0,8349	0,8436	0,8318
trainlm	0,9831	0,9657	0,9802

Az 5. táblázatban látható, hogy legnagyobb R^2 értéket a Levenberg-Marquardt legkisebb négyzetek módszerén alapuló tanítási algoritmus esetén tapasztaltam. A 6. táblázat az optimalizált hálózat R^2 és MAE értékeit láthatóak:

6. táblázat: Az optimalizált mesterséges neurális hálózat R^2 és MAE értékei

	R^2	MAE
Tanítás	0,9650	0,6616
Validálás	0,9470	0,7240
Tesztelés	0,9628	0,8089

A 6. táblázatban megfigyelhetők a különböző csoportok R^2 és MAE értékei. Megfigyelhető, hogy a legnagyobb R^2 értékek a tanítási, illetve tesztelési csoporthoz tartoznak. Érdeemes megjegyezni, hogy a R^2 és MAE értékek a többváltozós lineáris regresszióhoz (2. táblázat) képest sokkal kedvezőbbek, ami bizonyítja azt, hogy σ -momentumok és a hőmérsékletfüggő gáz-folyadék határfelületi feszültség között nem-lineáris összefüggés van. A következő ábrán (12. ábra) az optimalizált neurális hálózat segítségével becsült és a kísérletileg meghatározott gáz-folyadék határfelületi értékek összehasonlítása látható:



12. ábra: Az optimalizált neurális hálózat predikciós képességének bemutatása

A 12. ábrán az optimalizált hálózat számítási eredményei láthatóak. Az ábrán a folytonos fekete vonal a 45°-os egyenes, melynek pontjai az $R^2 = 1$ értékeket jelölik. Megfigyelhető, hogy ebben az esetben a 9. ábrához képest sokkal jobb becslést kaptunk a hőmérsékletfüggő gáz-folyadék határfelületi feszültségre. A 6. táblázat és a 12. ábra alapján kijelenthető, hogy az optimált neurális hálózat képes megfelelő pontosságú becslést adni a hőmérsékletfüggő gáz-folyadék határfelületi feszültségre. Ezt követően kiválasztottam a felhasznált molekuláris jellemzők (Klamt szett) közül azon elemeket, melyek legnagyobb mértékben határozzák meg a gáz-folyadék határfelületi feszültség

értékét. Ehhez a súlyok, perturbációs és COM eljárásokat alkalmaztam. Az egyes módszerek eredményeit a 7. táblázat tartalmazza:

7 táblázat: Az optimalizált meterséges neurális hálózat bemeneteinek érzékenységvizsgálata

Helyezés/Módszer	Súlyok	Perturbációs	COM
1	M_3	M_{HBacc3}	M_3
2	M_0	M_2	M_{HBacc3}
3	M_{HBacc3}	M_3	M_2
4	M_{HBdon3}	M_0	M_{HBdon3}
5	M_2	M_{HBdon3}	M_0

A 7. táblázat tárgyalásához elsőként be kell vezetnem az átlagos helyezés fogalmát. Az átlagos helyezés számításához az egyes momentumok különböző módszerek esetén elért helyezéseit átlagoltam. Az egyes momentumok átlagos helyezését a 8. táblázat tartalmazza:

8 táblázat: A különböző momentumok átlagos helyezése

Momentum	Átlagos helyezés
M_3	1,67
M_{HBacc3}	2
M_2	3,33
M_0	3,67
M_{HBdon3}	4,33

A 8. táblázatban látható a Klamt szett momentumainak átlagos helyezése. A különböző módszerekkel végzett vizsgálatok alapján kijelenthető, hogy az M_3 momentum által jelölt molekuláris tulajdonság befolyásolja legnagyobb mértékben a gáz-folyadék határfelületi feszültség értékét. Az M_3 momentum lényegében a σ -profil torzultsága, mely a molekula méretének hatását módosítja. Mivel a COSMO-RS elmélet a molekulák felületi töltéssűrűségének eloszlását veszi alapul, ezért az érintkező egységnyi felületelemekre jutó töltés mennyisége határozza meg a molekulák között kialakuló kölcsönhatás minőségét. Az M_3 momentum az érintkező felületegységek közötti elektrosztatikus kölcsönhatásokkal hozható kapcsolatba a molekula méretének módosításán keresztül, így nem meglepő, hogy ez a momentum befolyásolja

legnagyobb mértékben a felületi feszültség értékét. Ezt követi az M_{Hbacc3} momentum, mely a molekulák H-híd kötésre való hajlamát mutatja. Ebben az esetben nem lehet egyértelműen megállapítani ezen momentum fontosságát, ugyanis a bemeneti adatbázisnak csak egy része képes H-híd kötések létesítésére. A későbbi vizsgálatok során olyan szeparált bemeneti adatbázisokat fogok készíteni, melyek csak H-híd kötésre hajlamos vegyületeket fognak tartalmazni. Ezek alapján a fent említett jelenséget ki tudom szűrni. Az M_2 momentum a molekula elektrosztatikus árnyékoló hatása, mely gyakorlatilag az atomi szintű polarizációval van összefüggésben [48]. A dipol-dipol intermolekuláris kölcsönhatás mértéke arányos a polarizálhatósággal, ezért az M_2 momentum ezen kapcsolat kialakulásában játszik fontos szerepet. Az M_0 a molekula teljes felületét adja meg. Meglepő, hogy ez a momentum a vizsgálatok alapján nem játszik fontosabb szerepet, ugyanis a molekulák között kialakuló van der Waals-erők a molekulák felületének nagyságával arányosak. A H-híd donor tulajdonságokat leíró M_{Hbdon3} szintén nehezen értelmezhető a bemeneti adatbázis inhomogenitása miatt.

A vizsgálat eredménye alapján megállapítható, hogy a gáz-folyadék határfelületi feszültség molekuláris szintű értelmezése során az ún. Klamt-szett elemeit figyelembe véve az M_3 és M_{Hbacc3} momentumok játsszák a legfontosabb szerepet.

5. Összefoglalás

A munkám során gáz-folyadék határfelületi feszültség becslésére szolgáló modellt hoztam létre, melynek bemenő paraméterei a COSMO-RS elmélet alapján számítható σ -momentumok, illetve a hőmérséklet. Elsőként áttekintettem a felületi feszültség jelenségéhez és becsléséhez kapcsolódó szakirodalmi forrásokat, melyek alapján megállapítottam, hogy egy olyan modell, mely pusztán molekuláris jellemzők felhasználásával képes a gőz-folyadék határfelületi feszültség becslésére, igen fontos lehet a mérnöki gyakorlatban. A molekuláris jellemzőknek a kvantumkémiai COSMO számításra épülő COSMO-RS elméletből számítható σ -momentumokat választottam, melyek szakirodalma rövid bemutatásra került a dolgozban. A módszer bemeneti adatbázisának előállítás után többváltozós lineáris regressziót végeztem, mely során megállapítottam, hogy a σ -momentumok és a hőmérsékletfüggő felületi feszültség között nem-lineáris kapcsolat tételezhető fel. Ezért egy nem-lineáris modellezési technikával, a mesterséges neurális hálózatok segítségével hoztam létre a kapcsolat modelljét. Ehhez szükség volt a hálózat topológiájának optimalizálására, mely során

kiválasztottam a rejtett rétegek számát, a rejtett rétegbeli neuronok számát, a neuronok átviteli függvényét és a mesterséges neurális hálózat tanítási algoritmusát. Eredményként egy olyan modellt kaptam, mely megfelelően képes becsülni a hőmérsékletfüggő gáz-folyadék határfelületi feszültséget pusztán molekuláris információk felhasználásával. A dolgozat utolsó részében a létrehozott modell segítségével megpróbáltam molekuláris szinten magyarázatot adni a gáz-folyadék felületi feszültség kialakulására.

6. Irodalomjegyzék

1. Rideal, E.K., *Physical Chemistry of Surfaces*, 2nd ed., Arthur W. Adamson (Ed.). Interscience Publishers, New York (1967). Journal of Electroanalytical Chemistry and Interfacial Electrochemistry, 1968. **18**(4): p. 474.
2. Yanashima, R., et al., *Cutting a Drop of Water Pinned by Wire Loops Using a Superhydrophobic Surface and Knife*. PLOS One, 2012. **7**: p. 1-6.
3. Lee, B.-B., P. Ravindra, and E.-S. Chan, *New drop weight analysis for surface tension determination of liquids*. Colloids and Surfaces A: Physicochemical and Engineering Aspects, 2009. **332**(2–3): p. 112-120.
4. Lyman, W.J., D.H. Rosenblatt, and W.J. Reehl, *Handbook of Chemical Property Estimation Methods*. 1990, Washington DC: American Chemical Society.
5. Li, P., et al., *Estimations of surface tensions at different temperatures by a corresponding-states group-contribution method*. Fluid Phase Equilibria, 1996. **118**(1): p. 13-26.
6. Abbas, S. and S. Nordholm, *Simple Estimation of Bulk Thermodynamic Properties and Surface Tension of Polar Fluids*. Journal of Colloid and Interface Science, 1995. **174**(1): p. 264-267.
7. Romero-Martínez, A., A. Trejo, and F. Murrieta-Guevara, *Surface tension of isomers of pure hydrocarbons: a method for estimation and prediction*. Fluid Phase Equilibria, 2000. **171**(1–2): p. 1-10.
8. Macleod, D.B., *On a relation between surface tension and density*. Transactions of the Faraday Society, 1923. **19**(July): p. 38-41.
9. Sugden, S., *CXLI.-The influence of the orientation of surface molecules on the surface tension of pure liquids*. Journal of the Chemical Society, Transactions, 1924. **125**(0): p. 1167-1177.
10. Gardas, R.L. and J.A.P. Coutinho, *Applying a QSPR correlation to the prediction of surface tensions of ionic liquids*. Fluid Phase Equilibria, 2008. **265**(1–2): p. 57-65.
11. Khajeh, A. and H. Modarress, *QSPR prediction of surface tension of refrigerants from their molecular structures*. International Journal of Refrigeration, 2012. **35**(1): p. 150-159.
12. Veszprémi, T. and M. Fehér, *A kvantumkémia alapjai és alkalmazása*. 2002: Műszaki Könyvkiadó.
13. Klamt, A. and G. Schuurmann, *COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient*. Journal of the Chemical Society, Perkin Transactions 2, 1993(5): p. 799-805.
14. Klamt, A., *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*. 2005: Elsevier Science.
15. Klamt, A., *Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena*. The Journal of Physical Chemistry, 1995. **99**(7): p. 2224-2235.
16. Eckert, F. and A. Klamt, *Fast solvent screening via quantum chemistry: COSMO-RS approach*. AIChE Journal, 2002. **48**(2): p. 369-385.
17. Mehler, C., A. Klamt, and W. Peukert, *Use of COSMO-RS for the prediction of adsorption equilibria*. AIChE Journal, 2002. **48**(5): p. 1093-1099.
18. Winget, P., C.J. Cramer, and D.G. Truhlar, *Prediction of Soil Sorption Coefficients Using a Universal Solvation Model*. Environmental Science & Technology, 2000. **34**(22): p. 4733-4740.
19. Klamt, A., F. Eckert, and W. Arlt, *COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures*. Annual Review of Chemical and Biomolecular Engineering, 2010. **1**(1): p. 101-122.
20. Rarey, J., *Differences between COSMO-RS, COSMO-RS(OL) and COSMO-SAC*. 2006, DDBST GmbH.
21. Jasper, J.J., *The Surface Tension of Pure Liquid Compounds*. Journal of Physical and Chemical Reference Data, 1972. **1**(4): p. 841-1010.
22. Ahlrichs, R., et al., *Electronic structure calculations on workstation computers: The program system turbomole*. Chemical Physics Letters, 1989. **162**(3): p. 165-169.

23. Eckert, F. and A. Klamt *COSMOtherm, Version C2.1, Release 01.10.* 2009.
24. Xu, J., et al., *QSPR analysis of infinite dilution activity coefficients of chlorinated organic compounds in water.* Fluid Phase Equilibria, 2010. **291**(2): p. 111-116.
25. Katritzky, A.R., et al., *QSPR modeling of flash points: An update.* Journal of Molecular Graphics and Modelling, 2007. **26**(2): p. 529-536.
26. Guendouzi, A. and S.M. Mekelleche, *Prediction of the melting points of fatty acids from computed molecular descriptors: A quantitative structure–property relationship study.* Chemistry and Physics of Lipids, 2012. **165**(1): p. 1-6.
27. D’Archivio, A.A., et al., *Adsorption of s-triazines onto polybenzimidazole: A quantitative structure–property relationship investigation.* Analytica Chimica Acta, 2009. **650**(2): p. 175-182.
28. Járvas, G., C. Quellet, and A. Dallos, *Estimation of Hansen solubility parameters using multivariate nonlinear QSPR modeling with COSMO screening charge density moments.* Fluid Phase Equilibria, 2011. **309**(1): p. 8-14.
29. Lagarias, J.C., et al., *Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions.* SIAM J. on Optimization, 1998. **9**(1): p. 112-147.
30. Hansen, N., *The CMA Evolution Strategy: A Comparing Review*, in *Towards a New Evolutionary Computation*, J. Lozano, et al., Editors. 2006, Springer Berlin Heidelberg. p. 75-102.
31. Sun, S. and H. Liu, *6 - Particle Swarm Algorithm: Convergence and Applications*, in *Swarm Intelligence and Bio-inspired Computation*, X.-S. Yang, et al., Editors. 2013, Elsevier: Oxford. p. 137-168.
32. Katritzky, A.R., et al., *Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction.* Chemical Reviews, 2010. **110**(10): p. 5714-5789.
33. Baskin, I., V. Palyulin, and N. Zefirov, *Neural Networks in Building QSAR Models*, in *Artificial Neural Networks*, D. Livingstone, Editor. 2009, Humana Press. p. 133-154.
34. Guha, R. and P.C. Jurs, *Interpreting Computational Neural Network QSAR Models: A Measure of Descriptor Importance.* Journal of Chemical Information and Modeling, 2005. **45**(3): p. 800-806.
35. Guha, R., D.T. Stanton, and P.C. Jurs, *Interpreting Computational Neural Network Quantitative Structure–Activity Relationship Models: A Detailed Interpretation of the Weights and Biases.* Journal of Chemical Information and Modeling, 2005. **45**(4): p. 1109-1121.
36. Héberger, K., *Quantitative structure–(chromatographic) retention relationships.* Journal of Chromatography A, 2007. **1158**(1–2): p. 273-305.
37. Molnár, L., et al., *A neural network based prediction of octanol–water partition coefficients using atomic5 fragmental descriptors.* Bioorganic & Medicinal Chemistry Letters, 2004. **14**(4): p. 851-853.
38. Tompos, A., et al., *Predictive performance of “highly complex” artificial neural networks.* Applied Catalysis A: General, 2007. **324**(0): p. 90-93.
39. Gábor Horváth, M. Altrichter, and B. Pataki, *Neurális hálózatok.* 2006, Budapest: Hungarian Edition Panem Könyvkiadó Kft.
40. Chen, T., H. Chen, and R.-w. Liu, *A Constructive Proof and An Extension of Cybenko’s Approximation Theorem*, in *Computing Science and Statistics*, C. Page and R. LePage, Editors. 1992, Springer New York. p. 163-168.
41. Grippo, L. and S. Lucidi, *A globally convergent version of the Polak-Ribière conjugate gradient method.* Mathematical Programming, 1997. **78**(3): p. 375-391.
42. Fletcher, R. and C.M. Reeves, *Function minimization by conjugate gradients.* The Computer Journal, 1964. **7**(2): p. 149-154.
43. Powell, M.J.D., *Restart procedures for the conjugate gradient method.* Mathematical Programming, 1977. **12**(1): p. 241-254.
44. Marquardt, D., *An Algorithm for Least-Squares Estimation of Nonlinear Parameters.* SIAM Journal on Applied Mathematics, 1963. **11**(2): p. 431-441.
45. Garson, G.D., *Interpreting neural-network connection weights.* AI Expert, 1991. **6**(4): p. 46-51.

46. Dedecker, A., et al., *Application Of Artificial Neural Network Models To Analyse The Relationships Between Gammarus pulex L. (Crustacea, Amphipoda) And River Characteristics*. Environmental Monitoring and Assessment, 2005. **111**(1-3): p. 223-241.
47. Sung, A.H., *Ranking importance of input parameters of neural networks*. Expert Systems with Applications, 1998. **15**(3-4): p. 405-411.
48. Klamt, A. and F. Eckert, *COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids*. Fluid Phase Equilibria, 2000. **172**(1): p. 43-72.

7. Köszönetnyilvánítás

Köszönettel tartozom témavezetőimnek, dr. Dallos Andrásnak és dr. Járvás Gábornak a munkám során nyújtott sok segítségért, valamint szeretném megköszönni a Pannon Egyetem Fizikai Kémia Intézeti Tanszékén kutatómunkát végző munkatársaimnak, barátaimnak az inspiráló légkört, melyben munkámat végezhettem.

A kutatás a TÁMOP 4.2.4.A/2-11-1-2012-0001 azonosító számú „Nemzet Kiválóság Program - Hazai hallgatói, illetve kutatói személyi támogatást biztosító rendszer kidolgozása és működtetése konvergencia program” című kiemelt projekt keretében zajlott. A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.