BIOGEOKÉMIAI MODELLRE ÉPÜLŐ SZIMULÁCIÓS KERETRENDSZER FEJLESZTÉSE ÉS ALKALMAZÁSA

DOKTORI ÉRTEKEZÉS HOLLÓS ROLAND

Témavezetők:

dr. Barcza Zoltán egyetemi docens dr. Fodor Nándor tudományos főmunkatárs

Eötvös Loránd Tudományegyetem, Környezettudományi Doktori Iskola Környezeti Földtudomány Program

Doktori Iskola vezető: Dr. Turányi Tamás, egyetemi tanár



Eötvös Loránd Tudományegyetem, Természettudományi Kar Földrajz- és Földtudományi Intézet, Meteorológiai Tanszék DOI: 10.15476/ELTE.2024.009

Budapest 2024

Disszertációmat Márász Alice emlékének ajánlom, akinek a bátorítása nélkül nem kezdtem volna el a felsőfokú tanulmányaimat.

Tartalomjegyzék

1.	Bevez	etés	5
2.	Irodalı	ni áttekintés	
	2.1. A	talaj-növény rendszer folyamatai	8
	2.2. Fo	olyamatorientált modellek	11
	2.3. M	lodell-adat fúzió koncepció áttekintése	
	2.3.1.	Érzékenység-elemzés	14
	2.3.2.	Inverz modellezés	15
	2.4. K	eretrendszerek	23
3.	Adat é	s módszer	25
	3.1. Bi	iome-BGCMuSo	25
	3.2. A	Biome-BGCMuSo köré épülő keretrendszer: RBBGCMuso	
	3.2.1.	Az RBBGCMuso felépítése	
	3.2.2.	Az RBBGCMuso randomizációs algoritmusa	
	3.3. M	legfigyelési adatok	
	3.4. K	límaadatok	
	3.5. Ta	alajadatok	
	3.6. Pa	araméter optimalizáció	
	3.6.1.	A feltételes tartománycsökkentő módszer	
	3.7. A	modell térbeli kiterjesztését támogató keretrendszer: AgroMo	
	3.7.1.	AgroMo programozási háttere	
	3.7.2.	A modell térbeli kiterjesztésének lehetőségei	
	3.7.3.	AgroMo adatbázis és fájlrendszere	
	3.7.4.	AgroMo lekérdezésekhez használt metanyelv	
	3.7.5.	AgroMo szcenáriók és domének	
	3.7.6.	AgroMo kalibráció	
	3.7.7.	Disztribúció	
	3.8. K	límaváltozás hatástanulmány	
4.	Eredm	ények	61
	4.1. A	z RBBGCMuso keretrendszer	61
	4.1.1.	Ábrázolás, grafikon készítés	63
	4.1.2.	Lokális érzékenység-elemzés	64
	4.1.3.	Globális érzékenység-elemzés	64

4.2	2. Gl	LUE és CIRM	66
	4.2.1.	GLUE és utófeldolgozás a döntési fákon keresztül	66
	4.2.2.	Az iterációk eredményei és a döntési fák automatikus értelmezése	72
	4.2.3.	A modell teljesítményének számszerűsítése	76
4.3	3. Az	z AgroMo keretrendszer	78
	4.3.1.	AgroMo Site (cella-szintű futtatások)	79
	4.3.2.	AgroMo Plot (aggregálás és vizualizáció)	81
	4.3.3.	AgroMo Grid (térben explicit futtatások)	83
	4.3.4.	AgroMo Query (lekérdezések egyszerűsítése, riportkészítés)	85
	4.3.5.	AgroMo Map (térképek készítése)	86
	4.3.6.	AgroMo paraméter optimalizáció (ParAna)	88
4.4	4. Pr	ojekciók kukoricára	89
5.	Összef	oglalás	95
6.	Köszöi	netnyilvánítás	96
7.	Irodalo	mjegyzék	97
MEL	LÉKLE	Т	115

1. Bevezetés

A szárazföldi bioszféra, azon belül a növényzettel borított területek jelentős mértékben befolyásolják a légköri szén-dioxid (CO₂) mennyiségét (Canadell et al., 2021). Az ipari forradalom kezdete óta kibocsátott, antropogén eredetű CO₂ hatására a jelenlegi légköri koncentráció kb. 90 ppmmel¹ lenne több a szárazföldi területek folyamatos szénmegkötése nélkül (Friedlingstein et al., 2022). Ez azt jelenti, hogy a légköri CO₂ szén-egyenértékének megkötése a talaj-növény rendszerben egyfajta "ökoszisztéma-szolgáltatás", ami óriási értéket képvisel, és a jövőbeli alakulását kulcsfontosságú vizsgálni (Cox, 2019).

A szárazföldi növényzet nem csak a klímaszabályozásban betöltött szerepe miatt fontos, hanem meghatározó módon befolyásolja életünket az élelmiszertermelés, a nyersanyagellátás, a talajok egészségének fenntartása és sok más létfontosságú szolgáltatás révén. Emiatt a növényzet működésének vizsgálata sok évszázados múltra tekint vissza. Kiemelt helyen szerepelnek a konkrét gazdasági haszonnal járó területek, mint a növénytermesztés és az erdőgazdálkodás, illetve az állattenyésztés kapcsán a gyepgazdálkodás (Kipling et al., 2016; McDermid et al., 2014). Az erőfeszítések ellenére a talaj-növény rendszer működését meghatározó folyamatokat, illetve üvegházhatású gáz mérlegét jelenleg nem ismerjük kielégítő pontossággal.

A talaj-növény rendszer működését többek között matematikai modellek segítségével tudjuk vizsgálni és számszerűsíteni (Anderson et al., 2018; Jones et al., 2017). Ezek a modellek tipikusan valamilyen számítógépes szoftver formájában érhetők el. A talaj-növény rendszerrel kapcsolatos vizsgálatok eleinte kizárólag egy-egy részfolyamatot tekintettek át (pl. fotoszintézis, biomasszafelhalmozódás, evapotranszspiráció). Később az e folyamatok közötti kölcsönhatásokat leíró modellek is megjelentek, majd a még több folyamatot reprezentáló, almodulokból álló komplex modellek (talajmodellek, ökológiai és agrármodellek, globális növényzet-elterjedést és biogeokémiai folyamatait leíró modellek (Dynamic Global Vegetation Models; DGVM)), és később a föld-rendszermodellek (amelyek a klímamodelleket ötvözik a biogeokémiai modellekkel; Foley et al., 2000; Krinner et al., 2005; Lenton et al., 2006; Peng, 2000; Shevliakova et al., 2013). Ezen modellek almoduljainak az alapját a megfelelő szinten végzett terepi, illetve laboratóriumi kísérletek, megfigyelések adták. Az ember okozta klímaváltozással kapcsolatos vizsgálatok elindítása óta azonban egyre nagyobb területekre vonatkozóan, és hosszabb időskálán kellett a modellezésnek megtörténnie. A vizsgált rendszer országos, kontinentális, illetve globális léptékűvé vált, s ezen körülmények között kontrollált kísérletek végezni már nem lehet. Szerencsére a folyamatok közötti

¹ ppm: parts per million, azaz milliomod térfogatrész a száraz levegőre vonatkozóan. Megfelel a szén-dioxid és a száraz levegő mólarányának.

kölcsönhatások vizsgálatával és néhol egyszerűsítésekkel élve modellezni tudjuk ezeket a térben kiterjedt rendszereket is.

A modellek fejlesztése azonban számos új kihívás elé állította a kutatói közösséget. A jelenleg használt modellek rendkívül bonyolultak lettek, jelentős adatigénnyel rendelkeznek, számos olyan paramétert² használnak, amit a felhasználónak kell definiálnia, és számos kimenő adatot produkálnak (Bagnara et al., 2019). Emiatt a modellek alkalmazása a legelső lépéstől kezdve bonyolult és többlépcsős. Gazdasági és politikai vonatkozásai miatt a jelenleg használt modellek kapcsán óriási igény van arra, hogy azok használata egyszerűsödjön, és a modellezésben, illetve informatikában kevéssé jártas felhasználók is tudjanak értelmes és informatív szimulációkat végrehajtani.

Doktori munkám során olyan keretrendszereket hoztam létre, amelyek lehetővé tették egy komplex, integrált modellrendszerek alapelvei szerint megtervezett szoftver létrehozását. A rendszer magjaként használt modell néhány folyamata olyan empirikus összefüggéseket tartalmaz, amelyek paraméterértékei nem reprezentálnak fizikai folyamatokat. Mivel ezeket a paramétereket nem lehet a szakirodalom vagy mérések alapján meghatározni, ún. inverz módszerekkel kell beállítani őket. Ezek a modell paramétereit a szimulált folyamatok és azok mért megfelelői közti összefüggés alapján állítják be (és vele együtt választ adnak arra a kérdésre, hogy feltételezve a helyes modellműködést, mi lehet az a paraméteregyüttes, amellyel a modell lehető legjobban leírja a vizsgált rendszert). A módszer optimalizációs eljárásnak is tekinthető, mivel a modell és a mérések közötti hibát igyekszik minimalizálni. Ahhoz, hogy ezeket az inverz módszereket használni tudjuk, nagy mennyiségű és változatos adatforrásra lenne szükségünk. Amennyiben ezek megfelelő mennyiségben és minőségben rendelkezésre állnak, az inverziós modellezés hosszú múltja változatos lehetőségeket kínál (Tarantola, 2005). Azonban az esetek többségében vagy nincs elég megfigyelési adatunk, vagy ezek minősége nem megfelelő. Ilyen feltételek mellett kevésbé bízhatunk abban, hogy a modell a szimulált rendszerrel analóg módon működik, ami gátolja az alkalmazásának kereteit. Erre a limitációra ad hatékony megoldást az általam kifejlesztett modelloptimalizációs módszer, amely segítségével a vizsgált rendszerre vonatkozó tudományos előismereteink, más, önállóan elégtelen mennyiségű vagy minőségű adatforrás kombinálásával felhasználhatók az optimalizálásban. A dolgozatomban ezt a módszert, újdonságjellegénél fogva, szándékosan részletesebben és kiemeltebben mutatom be. Az optimalizált modell már alkalmas arra, hogy összetett, tudományos igényességű hatásvizsgálatot végezzünk vele. Ennek igazolására végeztem ensemble típusú hatásvizsgálatot a hazai kukoricatermelés jövőben várható változásának számszerűsítésére.

² Dolgozatomban a paraméter szó a modellek működését szabályozó változókat jelenti. Megkülönböztetendő a meghajtó meteorológiai adatoktól, illetve a kimenő, szimulált változóktól, amit a szakirodalom sok esetben szintén paraméterként említ.

Doktori munkám céljait a következő pontokban foglalom össze:

- Keretrendszer létrehozása a Biome-BGCMuSo biogeokémiai modellhez a benne rejlő lehetőségek kiaknázása céljából.
- Hiánypótló kalibrációs eljárás kifejlesztése folyamatorientált modellek számára gépi tanulás felhasználásával.
- Könnyen használható grafikus döntéstámogatói rendszer kialakítása a legkorszerűbb klímaprojekciók és más adatbázisok felhasználásával.
- A keretrendszer alkalmazása kukoricára, Magyarország területére, egy, az éghajlatváltozással összefüggő hatásvizsgálat céljából.

2. Irodalmi áttekintés

2.1. A talaj-növény rendszer folyamatai

A talaj-növény rendszer működését leíró matematikai modellek számszerűsítik a növényi, illetve talajban zajló folyamatokat, és azok kölcsönhatását. Ezek a modellek meghatározzák bizonyos tározók nagyságát (pl. talajban tárolt víz, talaj szerves anyag mennyisége, növényi részek biomasszája, levélzet tömege, haszonnövények esetén a szemtermés mennyisége stb.), illetve a légkör és az ökoszisztéma közötti anyagáramokat (pl. párolgás, szén-dioxid felvétel a légkörből, légzés stb.) (1. ábra). Tekintve a folyamatok sokrétűségét és komplexitását, a munka megköveteli, hogy egységesen definiált mennyiségeket használjon a kutatói közösség, és ezen mennyiségek összeegyeztethetők legyenek a kísérletek során megfigyelt által mennyiségekkel.



 ábra. A talaj-növény rendszer működésének egyszerűsített bemutatása. Az ábrán azon fő folyamatok vannak feltüntetve, amelyek a tipikus folyamatorientált modellekben megtalálhatóak. A kék színű nyilak a vízhez, a zöld a nitrogén vegyületekhez, a piros a hőáramhoz, a sárga az üvegházhatású gázokhoz, a barna pedig a talaj szervesanyag-tartalmához szorosan kapcsolódó folyamatokat ábrázolja.

Sajnos sok esetben a mai napig sincs egyetértés bizonyos mennyiségek jelentése kapcsán (pl. Lovett et al., 2006; Roxburgh et al., 2005). Tekintve, hogy a kor elvárásainak megfelelő modellek már az éghajlat szempontjából fontos üvegházhatású gázok mérlegét is megpróbálják

számszerűsíteni, a modellezett mennyiségek tárháza kibővült. Emiatt szükségszerű definiálni, és lehetőség szerint a szakirodalommal harmonizálni a származtatott mennyiségeket.

Az 1. táblázatban foglalom össze a talaj-növény rendszer működése szempontjából használt releváns mennyiségeket, és definiálom a jelentésüket. A biogeokémiai folyamatok kapcsán (ami jelen munka egyik központi eleme) Chapin et al. (2006) javasolt nevezéktant, amit felhasználtam a táblázat készítésekor.

Rövidítés	A mennyiség angol/magyar neve*	A mennyiség magyarázata
NPP	Net Primary Production/Nettó elsődleges produkció	A fotoszintézis és a növényi légzés különbsége, szén-egyenértékben vagy szárazanyagban kife- jezve. Adott időszakon belül az élő biomassza fel- halmozódásának a mérőszáma.
NEE	Net Ecosystem Exchange/ Nettó ökoszisztéma CO ₂ kicse- rélődés	A talaj-növény rendszer és a légkör közötti nettó CO ₂ áram. Az NEE eddy kovariancia mérések által származtatható.
GPP	Gross Primary Pro- duction/Bruttó elsődleges ter- mék	A fotoszintézis útján létrehozott anyagmennyiség. Általában szén-egyenértékben szokás kifejezni. A GPP az eddy kovariancia mérésekből fejlett techni- kával nagy pontossággal származtatható.
R _{eco} /TER	Total ecosystem respi- ration/ökoszisztéma szintű lég- zés	A talaj-növény rendszer légzése, azaz CO ₂ kibocsá- tása. A növényi komponens az autotróf respiráció, míg a talajok és az avar szervesanyag-tartalmának bomlásából származó CO ₂ a heterotróf respiráció.
ET	Evapotranspiration/ evapotranszspiráció	A talaj-növény rendszer vízmérlegét meghatározó egyik legfontosabb mennyiség. Több komponens összege (a talaj párolgása, a levélen keresztül tör- ténő transzspiráció és a növény által felfogott víz párolgása).
SWC	Soil water content/talajnedves- ség	A talajok pórusaiban tárolt víz mennyisége, általá- ban térfogatszázalékben kifejezve
LAI	Leaf area index/levélfelületi in- dex	Egységnyi felszínre vonatkozó egy oldalas levélfe- lület. Általában a teljes levélfelület fele.
Yield	Crop yield/hozam	A mezőgazdasági haszonnövény szemtermése egy- ségnyi területre vonatkozóan, a teljes nedves bio- masszára vonatkozóan (ami általában 14%)
DM	Grain dry matter/szem szára- zanyag tartalma	A szemtermés szárazanyag egyenértékben kife- jezve
HI	Harvest index/harvest index	A szemtermés és a talaj feletti teljes biomassza ará- nya

1. táblázat. A tala	j-növény rendszer	működését leíró	fontosabb men	nyiségek magy	/arázattal
			1		

* A magyar fordítás sok esetben nem egyértelmű, mert többfajta fordítás is előfordul a hazai szakirodalomban

A talaj-növény rendszer működését számos biotikus és abiotikus hatás befolyásolja. A modellek főleg az abiotikus hatásokat kezelik, de a szakirodalomban van példa kártevők, illetve betegségek hatásának leírására is (Jönsson et al., 2012). Az éghajlatváltozás kapcsán számos

tanulmány vizsgálta a szárazföldi bioszféra kapcsán és a változó környezeti tényezők hatását a növényi növekedésre és üvegházhatású gáz mérlegre (Albergel et al., 2018; Ciais et al., 2005; Reichstein et al., 2013; Teuling et al., 2010; Zscheischler et al., 2018). A jól ismert negatív környezeti hatásokon³ kívül (aszály, szélsőséges időjárás, hőhullám) fontos megemlíteni a légköri CO₂ koncentráció növekedésének pozitív hatásait is, ami a szakmában CO₂ trágyázás néven ismert (Ainsworth & Long, 2021). A folyamatorientált modellekben a környezeti hatások bemenő adatként határozzák meg a szimulált rendszer működését.

Tekintve a talaj-növény rendszerben zajló folyamatok meghatározó szerepét az éghajlati rendszerben, kiemelten fontos megérteni a változó környezeti állapot és a szárazföldi ökoszisztémák kölcsönhatását. A 2. ábra mutatja be a problémakör komplexitását.



2. ábra. A talaj-növény rendszer működésének kapcsolata az extrém időjárási eseményekkel, és a hatások jellege. A folytonos vonalak közvetlen hatásokat — pozitív hatást "+", a negatív hatást a "-" szimbólumok jelölik — míg a szaggatott vonalak indirekt hatásokat reprezentálnak. A vonalak vastagsága a fontosságot jelzi (a grafika a Reichstein et al. (2013) tanulmány első ábrájának a magyarítása).

Az ábrán látható, hogy a kölcsönhatások miatt a várható visszacsatolások intuitív módon nem származtathatóak, tehát a folyamatok számszerűsítése, a folyamatok közti kölcsönhatások leírása, a

³ Magyarországra, valamint általánosan az OECD államokra nézve

holisztikus szemlélet alkalmazása nélkülözhetetlen eleme a tudományos megismerésnek. A talajnövény rendszer működését változó éghajlati körülmények közt csak az ok-okozati összefüggések matematikai leírásával, azaz folyamatorientált modellek segítségével lehet leírni. Bár a modellek komplexitását növeli, az új keletű tudományos eredmények modellekbe történő implementálása ugyancsak meghatározó eleme a tudomány fejlődésének.

2.2. Folyamatorientált modellek

Dolgozatomban kizárólag folyamatorientált modellekkel foglalkozom, azon belül is a talajnövény rendszer működésének leírásával foglalkozó modellekkel. Itt két nagyobb modellcsalád létezik: a biogeokémiai modellek és a mezőgazdasági (crop) modellek, amelyek a jelen munka szempontjából egyaránt fontosak. Ezek a folyamatorientált modellek ismert biológiai, kémiai, növényélettani és fizikai folyamatokat írnak le, amelyek meghatározzák a talaj-növény rendszer működését, és általában az anyagmegmaradás elvére épülnek. A folyamatok leírása matematikai módon történik az egyes komponensek szinergikus leírásával, amely által új, magasabb szintű jelenségek is magyarázhatóvá válnak. Ezen modellek általában komplexebbek, mint az empirikus vagy statisztikai modellek, amelyek többnyire az adatokra jól illeszthető függvények segítségével származtatnak mennyiségeket, a folyamatok leírása nélkül (Di Paola et al., 2016). Azonban a folyamatorientált modellek bonyolultsága szükségszerű, hisz ilyen módon a modellek sokkal általánosabban alkalmazhatók változó környezeti feltételek mellett. A klímaváltozás hatásait, illetve a mitigációs lehetőségek keresésére almalmas modellek éppen ezért – legalább részben – folyamatorientáltak.

A folyamatorientált biogeokémiai modellek tárháza igen nagy. Ezek a modellek szimulálják az ökoszisztéma teljes szén- és vízmérlegét, és az utóbbi években szinte mindegyik kezeli a nitrogénmérleget is. Egyes modellek egyéb nyomelemeket is leírnak (foszfor, kálium). A CENTURY (Parton et al., 1987; Parton, 1996) és annak napi léptékű változata, a DAYCENT (Parton et al., 1998) a talajok széntartalmának és a tápanyagok, különösen a nitrogén és foszfor dinamikájának modellezésére összpontosítottak. A DNDC (Li, 1996) modell azokra a specifikus folyamatokra összpontosít, amelyek kritikusak a talajok nitrogén- és szénmérlegének szempontjából (nitrifikáció, denitrifikáció). Az ECOSYS (Metivier et al., 2009) és az IBIS (Foley et al., 1996) modellek a növények, a talaj és a légkör kölcsönhatásait is kezelik. Az olyan ismert modellek, mint például az LPJ (Sitch et al., 2000), a DALEC (Steinau et al., 2019), a Biome-BGC (Running & Coughlan, 1988), a JULES (Cox et al., 1999), a CLM (Oleson et al., 2008) és az ORCHIDEE (Krinner et al., 2005) képesek a teljes szárazföldi vegetáció működését jellemezni az ún. plant functional type (PFT) logika alkalmazásával (Bonan et al., 2002). A PFT-k definiálása segíti a modellező munkát oly módon, hogy megjelenés és funkcionalitás szerint csoportosítja a növényfajokat. Mezőgazdasági célú modellek esetén általában nincs szó a teljes szén- és nitrogén mérlegről, de jellegzetességük, hogy az egyéb tápanyag-limitációkat is leírják, és a termésbecslésben erősek. Ilyen modellek a teljesség igénye nélkül a CERES (Basso et al., 2016), az APSIM (Moeller et al., 2014), a STICS (Brisson et al., 2003), a CROPSYST (Stöckle et al., 2003), és a WOFOST (De Wit et al., 2019). Hidrológiai almoduljuk általában fejlett, ami felveti annak a lehetőségét, hogy bizonyos komponeseket átvegyenek a biogeokémiai modellek, mert azok sok esetben kevéssé hatékony talajnedvesség-sémával rendelkeznek (Sándor et al., 2016). Az APSIM és a STICS komplex modellezési rendszerek, amelyek számos haszonnövény típusra és a talajfolyamatokra összpontosítanak. A DSSAT rendszer az egyik legismertebb, döntéstámogatói célú modellező rendszer, ami egy olyan komplex eszköz, amely különféle növényfajokra specializált paraméterekkel támogatja a termelés optimalizálását (Zelenák et al., 2022). A DSSAT amiatt is említést érdemel, mert az annak részeként elérhető CERES modell a hazai fejlesztésű 4M modell alapját szolgáltatta (Fodor et al., 2014).

A folyamatorientált modellek alkalmazása nem egyszerű feladat (Bagnara et al., 2019). Komplexitásuk miatt nagy hangsúlyt kell kapnia a teszteknek és a független validációnak is. A modellezés azonban nem egy olyan folyamat, amely a modell beállításával és alkalmazásával véget ér. A térben és időben változó környezeti feltételeknek megfelelően a modell is folyamatos hangolást és fejlesztést igényel. A modell finomhangolását kizárólag a folyamatosan rendelkezésre álló adatok segítségével tudjuk elvégezni, tehát szükség van olyan módszerekre, amelyekkel a modelleket folyamatosan az adatokhoz tudjuk igazítani. A modellek térbeli alkalmazása is számos problémát vet fel. A következő fejezetben erről lesz szó.

2.3. Modell-adat fúzió koncepció áttekintése

A modell-adat fúzió (angolul "model-data fusion"; a továbbiakban MDF) koncepciója foglalja össze a biogeokémiai modellezés (és tágabb értelemben a földtudományokban használt modellezés) módszertanát (3. ábra; Williams et al., 2009).

Egy modell megfelelő alkalmazása ma már messze túlmutat a modell input igényének kielégítésén, a modell beállításán (szakirodalmi adatok, mérések, valamint becslések alapján) és futtatásán. A modellezési folyamat során szükséges megbizonyosodnunk arról, hogy a modell helyesen működik. Ehhez a modellt jellemezni kell: különböző beállításokkal kell futtatni, az eredmények konzisztenciáját kell ellenőrizni, továbbá mérési adatok alapján a kezdeti bizonytalanságot becsülni.



3. ábra. A modell-adat fúzió grafikus áttekintése (forrás: Williams et al., 2009).

Ennél a lépésnél még nagy szerepe lehet az ún. próba-értékelés módszernek (angolul "trial and error"), ahol a kutató az előzetes tudományos vagy gyakorlati ismeretei alapján ismeri meg a modell működését. Ez a módszer tulajdonképpen a gyakorlatban annyit jelent, hogy a kutató intuitív, de *nem megismételhető* és dokumentált módon változtat a modell paraméterein, majd lefuttatja a modellt, és vizsgálja, hogy mi változott. Jó esetben megfigyelési adatokkal is összeveti az eredményt; rosszabb esetben csak azt vizsgálja, hogy az előzetes elvárásoknak megfelel-e a modell eredménye (pl. reális-e a növény növekedése, levélfelületi indexe stb.).

Általában már az előzetes próbálgatások során látszik, hogy vannak olyan paraméterek, amelyeknek a változtatásai nem eredményeznek lényeges változást a vizsgált kimeneti változóban, és vannak olyanok, amelyek érzékenyek a változtatásra. Mivel a folyamatorientált modellek komplexek, a kölcsönhatások miatt sok paraméter egyidejű beállítására van szükség. A paraméterek beállítása előtt minden esetben értékelni kell a paraméterek beállításának költségét, és a modell predikciók pontosságának és precizitásának⁴ növekményét. Amennyiben a relatív költség magas, a paraméterbeállítás során nem jut kellő erőforrás a fontosabb, érzékeny paraméterek beállítására.

A próba–értékelés módszer korlátait túllépve lehetőségünk van korszerű matematikai elveken alapuló érzékenység-elemzést elvégezni. Egyszerűbb és matematikailag könnyebben kivitelezhető módon először egyes paraméterek értékeinek külön változtatásával lehet lokális érzékenységelemzést végezni, majd az ennél költségesebb globális érzékenységelemzést célszerű elvégezni, ahol a

⁴ Egy predikció precíz, ha kicsi a bizonytalansága; pontos, ha a prediktált értékek közel vannak a mért értékekhez.

paramétereket már nem egymástól függetlenül, hanem többdimenziós rendszerben *egyszerre* változtatjuk. Miután lecsökkentettük a beállítandó paramétereink számát az érzékenység-elemzés alapján, inverz modellezési technikákkal (más szóval modell-kalibrációval, vagy modell-optimalizációval) megbecsülhetjük a paraméterek bizonytalanságát. Ezzel a módszerrel a paraméterek legvalószínűbb értékét tudjuk becsülni valószínűségi kalibráció esetén, illetve a legkisebb hibát eredményező értékeit tudjuk meghatározni analitikus, heurisztikus vagy frekventista kalibráció esetén. Ezt követően a modell általánosítása (pl. hasonló ökoszisztéma modellezése kissé más talaj vagy meteorológiai adatok mellett), majd független adatokon történő validálása szükséges ahhoz, hogy megbizonyosodjunk arról, hogy a modell nem csak a tanulóhalmazon működik jól. A modell alkalmazására csak ezen vizsgálat után kerülhet sor. A MDF-ben a folyamatot minden esetben meg kell ismételni, ha megváltoznak a környezeti feltételek, a modellben verzióváltás történik, vagy új nagy mennyiségű adatra teszünk szert a vizsgált rendszerrel kapcsolatban. Ez óriási kihívást jelent a kutatói közösség számára.

A 3. ábra alapján fontos kiemelni a folyamat adatközpontúságát. Ez az a folyamat, ami összeköti a modellt a valósággal. A következő két alfejezetben a folyamat két legfontosabb részét, vagyis az érzékenységelemzést és az inverz modellezést mutatom be. Ez utóbbi a modell-adat-fúzió matematikailag legnehezebb feladata, kiváltképp, ha nem modellspecifikus, hanem általános megoldást szeretnénk nyújtani a problémára. Emiatt kiemelt figyelmet fordítok a problémakör felvázolására.

2.3.1. Érzékenység-elemzés

Az érzékenység-elemzésnek két célja van: megismerni/ellenőrizni a modell működését (a paramétereken keresztül), illetve eldönteni, hogy egy adott paraméter értékére egy adott helyen mennyire érzékeny a modell (vagyis egységnyi kis változásra mekkora kimeneti változással reagál). A leggyakrabban használt módszerek az adott pontbeli érzékenységet a modell, mint függvény differenciálhányadosaként definiálják (Saltelli et al., 2004). A deriváltat többnyire véges különbségekkel szokták közelíteni a szisztematikus paramétermintázás mellett, amely során a paraméter-intervallumok felosztását követően azt nézzük, hogy az egységnyi paraméter-változtatás mekkora változást eredményez a vizsgált eredmény vonatkozásában.

Egy alternatív módszer a variancia alapú érzékenység elemzés, ahol az adott paraméter hatását a modell varianciájára vonatkozóan a variancia dekompozíciójával határozzák meg (Saltelli et al., 2004). Ez az ún. Sobol-érzékenység módszer (Nossent et al., 2011), ami mára a legelterjedtebb érzékenység-elemzési eljárás. A módszert, mivel gyakran a modellre vonatkozó teljes valószínűségi leírás igen költséges és nem mindig lehetséges, a Monte Carlo módszerekkel szokták házasítani. A módszer hátránya, hogy amennyiben sok paraméterünk van, a Monte Carlo minták száma kis eséllyel lesz reprezentatív (a paraméter tér mérete a paraméterek számával exponenciálisan nő). Ezt a problémát segédfüggvények alkalmazásával lehet megoldani. Ekkor a segédfüggvényt a mintavételezett pontokra illesztve az már kellő módon reprezentálja a modellünket olyan módon, hogy a Sobol-indexek kiszámítása egyszerűbb, mint az eredeti modell esetében. Két segédfüggvényosztályt szoktak jellemzően alkalmazni: több-dimenziós egyeneseket (Verbeeck et al., 2006), valamint Gauß-folyamatokat (Rohmer & Foerster, 2011). Előbbi előnye az egyszerűsége és gyorsasága, hátránya viszont a linearitásra, kolinearitásra való érzékenysége.

A Biome-BGC modellcsalád kapcsán variancia-alapú érzékenységelemzést használt White et al. (2000), ami a mai napig a legelterjedtebb megközelítés (Miyauchi et al., 2019; Raj et al., 2014; Ren et al., 2022; Tatarinov & Cienciala, 2006)⁵

Habár az érzékenység-elemzési módszerek hasznos eszközök és a modell-adat fúzió egyik legfontosabb kezdeti részét képzik⁶, alkalmazásuk korlátait is fontos tisztán látnunk. Minthogy a különböző érzékenységelemzési módszerek más és más eredményre vezethetnek (Ren et al., 2022; Saltelli et al., 2004), a módszer kiválasztása és alkalmazása sohasem lehet teljesen objektív, így minden esetben a modellezett rendszer és a konkrét kutatási kérdések birtokában lehet őket alkalmazni.

2.3.2. Inverz modellezés

A folyamatorientált modellek sajátossága, hogy azokhoz jellemzően számos paraméter társul, amelyeket a felhasználónak kell beállítania minden modellezési feladat során. A munkám során felhasznált Biome-BGCMuSo modell esetén ez hatványozottan igaz, hiszen 197 ökofiziológiai paramétert használ, és 178 talajparamétert. Ezek közül persze számos paraméter könnyen beállítható a korábbi munkák, az alapértelmezett értékek (pl. Hidy et al., 2021), illetve a talajfájlok esetén a talajadatokra vonatkozó adatbázisok alapján, mint például a SoilGrids (Poggio et al., 2021).

A modellparaméterek beállítása, vagyis a paraméterezés összetett eljárás, mivel általában a modell állítható paramétereinek egy része nem mérhető közvetlenül (pl. valamilyen folyamat reprezentációjához kapcsolódó empirikus együtthatók). Egy másik probléma a modellparaméterek jól ismert bizonytalansága, különösen akkor, ha a modellt nagy térbeli léptékeken használják (Van

⁵ A Ren et al. (2022) tanulmány az Eredmények fejezetben leírt RBBGCMuso csomag érzékenység-elemző függvényét is használta.

⁶ Fontos megjegyezni, hogy az érzékenység-elemzés a modell-adat fúziótól függetlenül használható eszköz, amellyel egy előre meghatározott változócsoport viselkedését vizsgáljuk. A dolgozatban a módszert viszont kizárólag a változók számának csökkentésére használtam.

Oijen et al., 2005; Xiong et al., 2008; Angulo et al., 2013). A modellparaméterek beállításának a leginkább hatékony és objektív módja a fent említetteknek megfelelően az inverz módszerek alkalmazása, amit paraméter-becslésnek, paraméter optimalizálásnak, kalibrációnak, illetve modell optimalizálásnak is szokás nevezni (Braswell et al., 2005; Van Oijen et al., 2005; Sadegh and Vrugt 2013; Bilionis et al., 2015). A folyamat inverz modellezésnek tekinthető, mivel a modell kimenete és a kimenő adatokra vonatkozó mérési adatok segítségével határozzuk meg a paramétereket, azaz tulajdonképpen visszakövetkeztetünk a megfelelő paraméter értékekre.

Az inverz modellezés akkor és csak akkor tekinthető sikeresnek, ha a becsült paramétereken alapuló modell eredményei elfogadható mértékben közel állnak a "valóságot" reprezentáló megfigyelésekhez (némi, a mérési hibával számszerűsített zajjal). Ha a becslés bizonytalansága is alacsony, a modell alkalmasnak tekinthető a vizsgált rendszer pontos és precíz leírására. A kulcstényező itt a szimulációknak a megfigyelésekhez való "közelségének" és a paraméterbecslés bizonytalanságának a mérőszáma (ahol ez utóbbi valószínűségi eloszlással ábrázolható; Trudinger et al., 2007). A metrikát a célfüggvény (pl. valószínűségi módszerek esetében a likelihood), míg az utóbbit a bizonytalansági tartomány⁷ vagy eloszlás (pl. valószínűségi esetben a poszterior sűrűségfüggvény) számszerűsíti.

A közelmúltig a legtöbb inverz modellezési erőfeszítés csak a célfüggvényre (vagyis a modell jóságát leíró módszerre) irányult, a paraméterek bizonytalanságának becslése nélkül (Wallach et al., 2021). Gyakran alkalmazzák például az analitikai optimalizálási technikákat, mint pl. a Levenberg-Marquardt, vagy grádiens-ereszkedés módszerét, amelyek a hibát (például a mérési értékek modelltől vett átlagos négyzetes eltérését [MSE]) folytonosan differenciálható függvényként értelmezik, s ennek megfelelően klasszikus szélsőértékkeresési problémaként tekintenek a modell-kalibrációra. Ezek a technikák gyors konvergenciát biztosíthatnak alacsony stabilitás mellett (emiatt ideálisak a mélytanulási alkalmazásokhoz, ahol ezeket leggyakrabban használják). Az eljárás sikere a paraméterek kezdeti értékeitől és a metaparaméterek megválasztásától függ (pl. a tanulékonyságot jellemző arány (angolul "learning rate") a gradiens ereszkedésnél; Goodfellow et al., 2017; Tarantola, 2005). Összetett folyamatorientált modellek esetén a deriváltra vonatkozó információk általában nem állnak rendelkezésre, vagy a hibafüggvény nem differenciálható, ezért az ún. deriválásmentes módszerek az olyan modellek esetében, mint a Biome-BGCMuso előnyben részesülnek.

A deriválásmentes módszereken belül két alkategóriát különböztethetünk meg: a közvetlen keresésen alapuló módszereket és a valószínűségi módszereket. Az első csak az optimális paraméterértékek megtalálására alkalmas, és jobban ki van téve a "dimenzionalitás átkának"

⁷ A bizonytalanság itt a paraméterre vonatkozó tudásunk hiányára vonatkozik, amit sokféle módon fejezhetünk ki: bizonytalansági intervallummal, poszterior eloszlásfüggvénnyel, stb.

(Bellman, 1957)⁸, ami azt jelenti, hogy ezek a módszerek csak kis paraméterszám, azaz alacsony dimenziójú esetekben alkalmazhatók (Tarantola, 2005). Ha csak néhány paramétert kell optimalizálnunk, a szisztematikus keresés, vagyis a paraméter intervallumok diszkretizálásával egy minden lehetséges kombinációt kipróbáló szimuláció-sokaság végrehajtása gyors, egyszerű és stabil lehet. Ez azt jelenti, hogy az optimalizálás megismétlése után ugyanazokkal a hiperparaméterekkel (azaz az osztópontok számával) az eredmény közel lesz az eredeti optimalizálási kimenethez. Ezen módszer esetén a "dimenzionalitás átkát" jól szemlélteti, hogy ha például minden paraméter-intervallumot csak 10 részre osztunk fel, akkor n paraméter esetén 10ⁿ számú szimuláció elvégzésére van szükség, ami kívül eshet a számítástechnikai értelemben vett lehetőségeken. A paramétertér méretének exponenciális növekedésének kezelésére metaheurisztikus algoritmusok, például genetikai algoritmusok is alkalmazhatók, de általában a stabilitás vagy az eredmény egyértelműségének feláldozásával (Gogna és Tayal, 2013; White et al., 2022). A globális optimum megtalálása ilyen esetben sem garantált.

A deriváltmentes módszerek másik nagy ágát a valószínűségi módszerek képviselik, amelyek célja általában nem csak a paraméterek optimumának megtalálása, hanem az is, hogy ezekhez a paraméterekhez bizonytalanságot is szolgáltassanak (Van Oijen et al., 2005; Hartig et al., 2011). Ezen módszerek többnyire a paraméterekhez, vagy azok egy szűk intervallumához valószínűséget rendelnek, amely az adott szimuláció jóságával arányos. Erre leggyakrabban használt metrika a likelihood függvény, amely annak a valószínűségét adja meg, hogy adott paraméter értékeket választva (folytonos esetben annak egy kis környezetét) a mért értékeket adja a modell.

A likelihood függvény a hiba valószínűségi modellje. Mivel a centrális határeloszlás tétel következményeként a hiba gyakran normál eloszlást követ, leggyakrabban normál-eloszlást választanak likelihoodnak, amelyet a következő formula ír le:

$$\mathcal{L}(\theta|d\in\mathcal{D}) = \prod_{i=1}^{n_d} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\mathcal{M}(\theta)-d_i)^2}{2\sigma^2}}$$
(1)

ahol \mathcal{M} a modellt, θ annak egy paraméteregyüttesét, \mathcal{D} a mérési adatok halmazát, d_i egy adott mért értéket, σ a mérés szórását, míg n_d a mérési adatpontok számát jelöli.

Ezek a módszerek rugalmasak⁹, és természetesen lehetőséget biztosítanak a paraméterek kalibrálására a modellezési, paraméterezési és mérési hibák egyidejű figyelembevétele mellett (Tarantola, 2005). A rugalmasságnak azonban ára van. Minél rugalmasabb a rendszer, annál több

⁸ A lehetséges paraméterkombinációk száma (paramétertér mérete) a paraméterek számának növelésével exponenciálisan nő.

⁹ Rugalmasak abban az értelemben, hogy szélesebb az alkalmazhatósági körük, és nagyobb finomhangolásra van lehetőség a kalibrációra vonatkozóan.

hiperparamétert és feltételezést kell alkalmazni a munkafolyamat egyszerűsítése érdekében. Egy valószínűségi modell felépítése általában sokkal nehezebb, mint egy jó célfüggvény megtalálása.

A modellek paramétereinek becslésére a legegyszerűbb valószínűségi módszer a maximum likelihood (ML) becslés, amely használható egyszerűbb és bonyolultabb modellek esetében is. Az előbbi esetben a likelihood függvény maximuma akár zárt alakban, analitikus módon is meghatározható. Bonyolultabb esetekben a likelihood függvény maximumát mintavételezéssel, Monte Carlo módszerrel közelítjük. Az eljárás lényege, hogy a paramétertér véletlen mintavételezését követően a likelihood függvényt kiértékeljük (vagyis megvizsgáljuk, hogy milyen közel van adott paraméter esetén a modell kimenete a megfigyelésekhez). Azt a paraméterkombinációt tekintjük optimálisnak, amelyik a legkisebb hibát eredményezte, vagyis amelyhez a legnagyobb likelihood érték tartozik. Már a módszer rövid leírásából is egyértelműen látszik, hogy a mintavételezési eljárásnak és a mintavételezett pontok számának kiemelkedő szerepe van az eljárás alkalmazhatósága szempontjából. A mintavételi algoritmus akkor tekinthető optimálisnak, ha minimális számú mintával a likelihood függvény és a likelihood függvény mintázásából származó empirikus eloszlásfüggvény közti eltérés minimális. A mintavételi eljárás hatékonyságától függően a "dimenzionalitás átka" itt is jelentős problémát jelenthet. A legnagyobb probléma azonban az, hogy ha a mintavételezett paraméterek száma alacsony, valamint, ha kevés a mérési adat, akkor az ML-érték nagyban függ az aktuálisan mintázott paraméterektől, s újabb mintázás esetén más ML-hely és -érték lesz az eredmény¹⁰.

Ennek a problémának a kezelésére egy lehetséges megoldást nyújt az ún. GLUE (Generalized Likelihood Uncertainty Estimation) módszer, amey nem az összes likelihood értéket használja a paraméterek bizonytalanságának becslésére, valamint az optimális paraméter meghatározására, hanem csak az úgynevezett "jól viselkedő" értékeket, amelyeket általában a mintavételezett paraméterek felső *x* percentiliseként (jellemzően *x*=5) határoznak meg (Beven & Binley, 2014; Prihodko et al., 2008; Sexton et al., 2016). A "jól viselkedő" paraméterértékek mindegyike "jó" paraméterértéknek minősül. Közülük az ML értékkel rendelkező paraméterérték csak egy a sok közül, és nincs bizonyítékunk rá, hogy kisebb validációs hiba jellemzi, mint a többit. A módszer a "jól viselkedő" paraméterek eloszlásának (általában) 2,5., illetve 97,5. percentilise közti tartományt tekintik a likelihood bizonytalansági tartományának¹¹. Ebből a tartományból véletlenszerűen mintázva, a hasonlóan jó parametrizációk között nem szükséges döntenünk, ráadásul az ezekkel a

¹⁰ A gépi tanulásban ez az ún. Bias-Variance Tradeoff jelenség. Az alacsony szisztematikus hiba (bias) ára gyakran a nagy variabilitás.

¹¹ Ez az intervallum nem tekinthető sem konfidencia, sem ún- "credible" intervallumnak, ahol egy adott valószínűséghez tartozó credible intervallum az a tartomány, amelyben a paraméter az adott valószínűségben megtalálható.

paraméterkombinációkkal futtatott modellek kimeneti értékeinek a szórása jól reprezentálja a modellezés bizonytalanságát, az ensemble átlaga pedig egy stabil becslést adhat a modellezett értékekre.

A valószínűségi módszerek másik nagy családja a Bayes-módszereké. Ezek a módszerek általában rugalmasabbak, mint a frekventista megközelítések (például ML), és lehetőséget biztosítanak a paramétertérrel kapcsolatos korábbi ismeretek (újra)hasznosítására. A korábbi ismeretek felhasználása megakadályozza, hogy az optimalizáló algoritmusok túlságosan "magabiztosak" legyenek a kis mennyiségű adatponton alapuló becslésben. A Bayes-szabály az az egységesítő eszköz, amellyel a paraméterekre vonatkozó előzetes ismereteink a likelihood függvény segítségével frissíthetőek. Az így keletkező ismereteket a paraméterek valószínűségeiről poszterior sűrűségfüggvénynek hívjuk (Gelman et al., 1995). Ez a sűrűségfüggvény valószínűségi információt szolgáltat számunkra a paramétereinkről. Például olyan kérdésekre adhatunk választ, hogy mekkora a valószínűsége annak, hogy egy adott paraméterérték egy bizonyos tartományon belül jelen van? Ezen információk segítségével megbecsülhetünk egy paramétert annak karakterisztikus pontjaival, mint például a maximális poszterior értékkel (MAP), miközben információt adhatunk a bizonytalansági intervallumokról, például a legmagasabb poszterior sűrűségű intervallumról (Highest Posterior Density Interval, HPDI). Ha azonban a módszertől csak azt várjuk el, hogy megbízható előrejelzést készítsen, a MAP-érték használata helyett a Bayes-modellátlagolás (BMA) lehet a legjobb megoldás (Hinne et al., 2020), mivel az egyszerűbb ensemble-módszerrel szemben a poszterior eloszlás összes információját felhasználja a kimenet előrejelzéséhez. Továbbá a poszterior eloszlás közvetlenül felhasználható hipotézisvizsgálathoz például a Bayes-faktorok (BF) segítségével (Gelman et al., 1995).

Minden erősségük ellenére — ha megfelelően hajtják végre őket — a Bayes-módszerek kétségkívül bonyolultak. Így az inverz modellezési folyamat során túlzottan sok feltételezést és számos döntést kell hozni. Először is, megfelelő likelihood és prior eloszlásfüggvényeket kell választani. A választás mindig a problémához kapcsolódik, de nem mindig nyilvánvaló. Például nem létezik általánosan alkalmazható, nem informatív a priori eloszlás (Pericchi és Walley, 1991), bár sokan az egyszerűsége miatt egyenletes eloszlást használnak erre a célra (Pericchi és Walley, 1991; Gelman, 1996; Gelman és Yao, 2020; Wallach et al., 2021). Továbbá, annak ellenére, hogy a megfelelő likekilhood függvény kiválasztása az egyik legfontosabb döntés, amit meg kell hozni (Trudinger et al., 2007; Dumont et al., 2014), a modellezők általában még mindig kizárólag a normál likelihood függvényt használnak (Wallach et al., 2021). A valószínűségi függvénycsalád kiválasztása azonban még mindig nem elegendő a megfelelő priorok vagy a globális likelihood függvény meghatározásához. Számos további feltételezéssel kell élnünk, mint például a mérések függetlensége, valamint a mérési és a modellhibák függetlensége (Tarantola, 2005). Ezek közül messze a

leggyakoribb feltételezés az, hogy csupán a marginális poszterior függvények vizsgálatával következtethetünk a poszterior függvényre. A marginális eloszlásfüggvények függetlenségére vonatkozó feltételek azonban sok esetben nem teljesülnek (ld. 4. ábra). Ilyen esetek kezelésére ez idáig csupán kevés kísérlet született, és ezek is általában csak páronként vagy egyes esetekben 3D-s szórásdiagramokkal vizsgálják a paraméterek közötti összefüggéseket (Sadegh és Vrugt, 2013; Beven és Binley, 2014; Her és Chaubey, 2015), mivel ennél több paraméter egyidejű figyelembevételére ezek a módszerek nem képesek. A mintavételi eljárásra vonatkozó feltételezések hasonlóak: a paraméterek javasolt értékeinek többségét egymástól függetlenül mintavételezzük.



4. ábra. Monte Carlo mintázás utáni likelihood eredmények ábrázolása két paraméter esetén (bal és jobb oldali ábra). Ezen marginális eloszlások látszólag függetlenek egymástól, de ha a második paraméter medián értékénél lévő pontokat vizsgáljuk, egyértelmű mintázat látható a likelihood-értékekben (középső ábra, piros színű pontok).

A paramétertér dimenzionalitásától függően a valószínűségi kalibráció számításigényes. Ezért előnyösebbek azok a módszerek, amelyek kevesebb paraméter optimalizálásával is informatív eredményeket tudnak nyújtani. Ha a bemeneti paraméterek függését figyelembe vesszük, kevesebb szimulációra van szükség (Bloom és Williams, 2015). Ha például a modellnek két bemeneti paramétere van, és a paraméterek összege 1 kell legyen, akkor a feltételnek megfelelő paraméter véletlenszerű egyenletes mintavétellel történő megkeresésének valószínűsége gyakorlatilag 0. Mivel a konvex összefüggések a legelterjedtebbek, a kifejezetten konvex esetekre tervezett mintavételi módszer alkalmazása pozitív hatással lehet az eljárás hatékonyságára. Egyébként minden konkáv régió felosztható diszjunkt konvex régiókra, ahol a bemeneti paraméterek egymástól függetlenül mintavételezhetők¹². Ha a konvex régió nem politóp¹³, akkor több politóppal közelíthető. Ezért a konvex politópos régiók mintavételezésére használt technikák elég általánosak ahhoz, hogy jó

¹² Például ha két paraméter lehetséges kombinációi egy hatágú csillag belső pontjai, a csillag felbontható 7 konvex tartományra: a hat ága hat háromszög, valamint a középső hatszög.

¹³ A politóp a sokszögek több dimenziós általánosíta. Jelen kontextusban a paramétertér minden esetben egy politóp, vagyis véges több dimenziós objektum. Három dimenzió esetén pl. egy téglatest vagy egy tetraéder.

jelöltek legyenek bármely paraméteres régió mintavételezésére. Az egyik ilyen algoritmus a Hit and Run mintavételezés (Chen and Schmeiser, 1996), amelyet (legjobb tudásunk szerint) még nem alkalmaztak folyamatorientált modellekre.

Ezen a ponton fontos megjegyezni, hogy még mindig nyitott kérdés, hogy a Bayes-módszer rendelkezik-e előnyökkel a frekventista módszerrel szemben, ha a Bayes-módszereket nem megfelelően alkalmazzák (Gelman és Yao, 2020). Számos tanulmány foglalkozik a különböző módszerek gyakorlati alkalmazásával, de nagyon messze vagyunk attól, hogy "legjobb gyakorlatokat" vagy általánosan alkalmazható megoldásokat kínáljunk a modellező közösségnek.

Vannak kísérletek az optimalizálás sikeresebbé tételére. A modellek összetettsége és ezzel arányosan a paramétertér növekvő mérete miatt egyre több megfigyelésre van szükség ahhoz, hogy az inverz modellezést magabiztosan lehessen elvégezni, és a probléma legtöbbször még mindig alulhatározottnak tekinthető (Sadegh és Vrugt, 2013). A modellező az alulhatározottság problémáját a rendszerre vonatkozó meglévő tudásának a kalibrációs folyamatba történő integrálásával tudja kezelni. A bemeneti paraméterek kapcsolatára vonatkozó információk megadása egy nemrégiben javasolt új irányzat (Richardson et al., 2010; Bloom és Williams, 2015). Ez a megközelítés azon a felismerésen alapul, hogy a szakértői tudás alapján nem minden paraméterkombináció valósághű.

A bemeneti paramétertéren alkalmazott megkötések módszere nem az egyetlen lehetőség a modelloptimalizálás eredményeinek javítására. A folyamatorientált modellek természetüknél fogva összetettek, és egynél több kimeneti folyamot eredményeznek. A modellek kalibrálása ezek közül csak az egyiket figyelembe véve "rossz okokból jó eredményeket" eredményezhet (Beven és Binley, 2014). Ennek következtében az így kapott paraméterezéssel a modell a tanuló-adatokon meglehetősen jól prediktálja az adott változót, de a validálási adatokon kudarcot vallhat (ez a túlillesztés tipikus esete). A túlillesztés esélye akkor a legnagyobb, ha a paraméterekhez ún. ekvifinalitás társul, ami azt jelenti, hogy az előre meghatározott intervallumban a különböző paraméterértékek valószínűsége hasonló (5. ábra).

Ekvifinalitás esetén nem tudunk objektíven választani a posterior paraméterek közül, ami azt jelenti, hogy a paraméterek bizonytalanságát nem lehet csökkenteni, azaz az intervallumot szűkíteni (Beven és Freer, 2001). Az ekvifinalitás nem feltétlenül jelenti a szimuláció kudarcát, hanem inkább az inverziós módszer korlátja.



5. ábra. Példa az ekvifinalitásra. Mindkét ábra egy biogeokémiai modell optimalizálása során Monte Carlo módszerrel végzett kalibráció pontdiagramját mutatja (a GLUE nómenklatúra szerint az ábratípust "dotty plot"-ként szokás említeni). Az x tengelyen a paraméter előre definiált intervalluma látható ([0,58; 0,73]), míg az y tengely mutatja a modell jóságát (likelihood értékek). Minden pont egy Monte Carlo módszerrel randomizált paraméter alapú szimulációnak felel meg. A bal oldalon egy jól beállítható paraméter, míg a jobb oldalon egy nem jól beállítható paraméter látható, ahol utóbbi esetben fennáll az ekvifinalitás. Ez azt jelenti, hogy a randomizáció során egyaránt előfordultak jó és rossz szimulációk is.

Az ekvifinalitás elkerülésének egyik lehetséges megoldása a multi-objektív¹⁴ kalibráció alkalmazása. Több megfigyelési adatfolyamra alapozott több objektív függvény alkalmazásával az eredmény stabilabb lehet, és kevesebb paraméterhez fog kapcsolódni ekvifinalitás (Her és Seong, 2018). A multi-objektív kalibráláshoz azonban jó minőségű megfigyelési adatokra van szükség a sikeres kalibráció biztosításához. A kimeneti adatfolyamok egymástól való függése miatt egy újabb adatfolyam hozzáadása a folyamathoz bonyolult munkafolyamathoz vezethet. A több adatfolyamból származó zaj pedig sokszor nem additív (vö. hiba-kovariancia mátrix). A valós életben általában nincs elég adatunk a multi-objektív kalibráláshoz, vagy az adatok túl zajosak.

A multi-objektív kalibrációval együtt a modellező alkalmazhat néhány visszautasítási szűrési (angolul "rejection filtering"; innentől az angol kifejezést használom) technikát. E megközelítések egyik fő előnye a modell kimeneteinek különböző statisztikái közötti korrelációkkal szembeni ellenálló képesség (Hartig et al., 2011). A "rejection filtering" alkalmazása önmagában azonban nem zsugorítja a paraméterteret, valamint ezek a módszereknehezen interpretálhatók. Ez az egyik oka a hagyományos (implicit) "rejection-filtering" ezért a hagyományosan (implicit) "rejection-filtering" alapú "trial-and-error" megközelítést alkalmazó modellezők nem rendelkeznek elegendő kontrollal az inverz modellezési eljárás felett. Továbbá, ha a szűrt, elfogadható szimulációk aránya túl alacsony,

¹⁴ Magyarul többcélú-optimalizálás. Több szempontot egyidejű figyelembevételével történő paraméterbeállítást jelent, amely során a hibafüggvényt minimalizáljuk.

az utólagos mintavételi módszerek konvergenciája különösen lassú vagy akár megkérdőjelezhető is lehet.

A valószínűségi módszerek végrehajtásának bonyolultsága és a használhatatlan eredmények (pl. az ekvifinalitás) fennállása lehet a fő oka annak, hogy a kutatók többsége még mindig a fent már említett "trial-and-error" módszert részesíti előnyben a modell-optimalizálásban (Wallach et al., 2021). Nyilvánvalóan nagy szükség van olyan értelmezhető, könnyen használható és hatékony módszerek kidolgozására, amelyek támogatják a modellezőket a modellek optimalizálásának javításában, különösen nagy dimenziójú esetekben (pl. komplex biogeokémiai modellek esetén, mint a Biome-BGCMuSo).

2.4. Keretrendszerek

A modell-adat fúzió részleges vagy teljes megvalósítása óriási feladat, és általában feltételezi egy nagyobb, stabil finanszírozású kutatócsoport meglétét. Minden komponens módszertani problémákkal terhelt, és igényli a mérési adatok rendelkezésre állását, lehetőleg több adattípusra (NEE, GPP, LAI, SWC, termésmennyiség, biomassza; 1. táblázat). Ez utóbbi esetén kiemelendő, hogy nem csak pontbeli mérési adatokra támaszkodunk, hanem a légköri koncentráció-mérések, illetve műholdas távérzékelésen alapuló adatokat is széles körben használnak a modellek optimalizálására, illetve validálására (Ciais et al., 2014). Tekintve a folyamatorientált modellek nagy számát és igen eltérő architektúráját, nem létezik univerzális megoldás az MDF kivitelezésére. Az egyes modellek köré diverz és egyedi megoldások születtek, amelyek az MDF bizonyos komponenseit képesek megvalósítani.

Az egyik ilyen ismert csomag a Predictive Ecosystem Analyzer (PEcAn) (LeBauer et al., 2013), amely egy regionális skálájú adatasszimilációs rendszer. Több biogeokémiai modell alkalmazását támogatja (BioCro, CLM, DALEC, ED2, GDAY, SIPNET stb.), és egyaránt alkalmas érzékenység-elemzésre, bizonytalanság becslésre, állapotbecslésre és paraméter optimalizálásra Bayes-módszerrel.

Egy másik fontos keretrendszer az LPJ-GUESS R csomag (rLPJGUESS), amlyik az ismert LPJ modellt alkalmazza (Bagnara et al., 2019). A csomag támogatja az érzékenység elemzést, a Monte Carlo alapú paraméter optimalizálást, és több klímaszcenárió alapján a modell térbeli futtatását.

A talajokban zajló folyamatok általános célú leírására született a SOILR szoftvercsomag (Sierra et al., 2012). A csomag érdekessége, hogy általánosítja a talaj szervesanyag bomlás folyamatait leíró modelleket, és egy magasabb szintű szimulációs platformot kínál. Az egyedi fejlesztésű ismert talajmodellek (mint pl. RothC) is rekonstruálhatók az általános struktúra alapján.

Említést érdemelnek még az evapotranszspiráció számítására létrehozott R csomagok (De Oliveira et al., 2019; Guo et al., 2016), illetve a NUCOMBog (Pullens et al., 2017), amely vizenyős területek biogeokémiai modellezésére alkalmas.

A leginkább összetett modellezési keretrendszerek az ún. Integrált Modellek (Ewert et al., 2015). Az IM-ek célja, hogy más tudományterületeket is bekapcsoljanak a modellezési folyamatba (pl. közgazdaság, kockázatelemzés, környezetvédelmi szempontok, egészségügy) oly módon, hogy a döntéshozatalban is közvetlenül használható eszközöket kapjunk, megőrizve a tudományos alaposságot és precizitást. Az IM-ek a mai napig gyerekcipőben járnak, használatuk pedig nem terjedt el olyan mértékben, ahogy arra számítani lehetett pár évvel ezelőtt.

3. Adat és módszer

3.1. Biome-BGCMuSo

A disszertációban bemutatott munka a Biome-BGCMuSo nevű folyamatorientált biogeokémiai modell köré épül. Emellett a bemutatott RBBGCMuso, illetve AgroMo szoftverek is a Biome-BGCMuSo modellen alapulnak. Ez konkrétan azt jelenti, hogy a modell futtatása és paraméterezése a központi probléma, amellyel foglalkozom, és emellett keretrendszereket építek ki a modell funkcionalitásának minél szélesebb körű kihasználása céljából.

A Biome-BGC egy általános célú biogeokémiai modell, amely a szárazföldi ökoszisztémák teljes szén-, nitrogén- és vízháztartását szimulálja (Churkina et al., 2003; Di Vittorio et al., 2010; Running & Coughlan, 1988; Bond-Lamberty et al., 2005) Magyarországon Hidy Dóra és Barcza Zoltán adaptálták a Biome-BGC 4.1-es verzióját (Hidy et al., 2006), és felismerve a modell korlátait kezdték el fejleszteni Biome-BGCMuSo néven. Ily módon a Biome-BGCMuSo az eredeti Biome-BGC modell egyik ága vagy módosulása (Hidy et al., 2012, 2016, 2021, 2022).

Az immár közel 20 éve zajló fejlesztések során az eredetihez képest számos tekintetben jelentősen javították és bővítették a Biome-BGC funkcionalitását. A fejlesztések a talajfolyamatokkal, az emberi beavatkozást leíró folyamatok leírásával, a növényi stressz számszerűsítésével és számos más folyamattal foglalkoztak. A modellfejlesztés egyik fontos mérföldköve volt egy 10 rétegű talaj¹⁵ almodul létrehozása kifinomult talajnedvesség mérleg rutinnal, a talajon belüli C- és N-dinamika rétegenkénti leírásával, valamint a részletes nitrifikációs/denitrifikációs rutin megvalósításával. A különböző stressztényezők (szárazságstressz, nitrogénstressz, hőstressz) javítása szintén jelentős előrelépés volt. Hőfok-nap (angolul "growing-degree day") alapú, fenofázis-specifikus allokációs rutin is bekerült, amely lehetővé tette a növények fejlődésének részletesebb és pontosabb szimulációját. Mezőgazdasági vonatkozású fejlesztések kapcsán beépítésre került a hazai 4M modellből (Fodor, 2006) több, mezőgazdasági haszonnövényekre jellemző mechanizmus (pl. virágzáskori hőstressz, vernalizáció, csírázás stb.). A szántóföldi szimulációkban a megfelelő eredményekhez elengedhetetlenek a részletes gazdálkodási információk megadása (beleértve a kijuttatott műtrágya időzítését és mennyiségét, a vetés időpontját, a talajművelés típusát és időpontját, a betakarítás időpontját, a mezőgazdasági melléktermékkel való gazdálkodás módját). A jelenlegi Biome-BGCMuSo modell kombinált biogeokémiai-növénytermesztési modellnek tekinthető, amely mindkét tudományterületen a legkorszerűbb folyamatok reprezentációjával rendelkezik (Hidy et al., 2022). A modell ún. "algorithm ensemble" opcióval is rendelkezik, ami azt jelenti, hogy bizonyos

¹⁵ Ezek a rétegek csak a modell konktextusában értelmezhetőek, és a rétegek homogénnek tekintendők.

folyamatok (mint pl. a talajnedvesség vertikális mozgása a talajban, a fotoszintézis, a növényi stressz leírása) két különböző algoritmus szerint is modellezhető, így vizsgálható a modell-logika hatása pl. a paraméterbecslésre vagy a modell működésére. A modell alkotói meghatározó fontosságúnak tartották a szoftver részletes dokumentációját is, felhasználói kézikönyv formájában (Hidy et al., 2021).

A modell sok évet átölelő fejlesztése során én is bekacsolódtam a munkába. Hozzájárultam többek között a fejlettebb hibakezelés kialakításához, a forráskódban lévő hibák javításához, a kényelmi funkciók bevezetéséhez stb.

Ebben a tanulmányban a Biome-BGCMuSo v6.3 verziót használtam. A Biome-BGCMuSo paraméterezése nem csak az állítható ökofiziológiai paraméterek nagy száma miatt összetett feladat, hanem azért is, mert a sikeres szimulációhoz bizonyos szabályoknak (azaz kényszerek) kell megfelelniük (Hidy et al., 2021). Egyes paraméterek esetében ezek a szabályok viszonylag egyszerűek (pl. az avar C:N arányának nagyobbnak kell lennie, mint a levelek C:N arányának), más esetekben viszont viszonylag összetettek (pl. a szén különböző növényi tározók közötti elosztását szabályozó paraméterek összege 1). Ezek a szabályok megnehezítik a modell optimalizálását, mivel a Monte Carlo keretrendszerben a modellt vezérlő kiválasztott paraméterek előre meghatározott intervallumain belül véletlen számokat generálnak (ami n változó esetén egy n dimenziós paraméterek esetében a véletlen számok összege az esetek többségében nem lesz 1, ami nagyszámú sikertelen szimulációt eredményez.

A Biome-BGCMuSo C programozási nyelvben íródott és a forráskódja szabadon hozzáférhető, részletes dokumentációval¹⁶ (Hidy et al., 2021). A modell futtatási környezete a hagyományos DOS ablakra épül Windows alatt (és parancssoros módra UNIX/Linux alatt), ami nem tekinthető korszerű megoldásnak. A bemenő adatokat szöveges (ASCII text) formátumban kell megadni, beleértve a meteorológiai bemenő adatokat, a modell beállításait tartalmazó ún. INI fájlokat, az emberi beavatkozást leíró menedzsment fájlokat, a talajfájlokat (SOIL), illetve ökofiziológiai fájlt (EPC) (6. ábra).

A Biome-BGC zsargon szerint a modellezett rendszer kezdeti állapotát az ún. spinup szimulációval állítjuk be, aminek a beállításait hagyományosan a spinup.ini vagy ahhoz hasonló beszédes nevű INI fájllal végezzük. A spinup vége egy olyan, az adott hely éghajlatával egyensúlyban lévő rendszer, amellyel már lehet végezni a felhasználó igényei szerinti vizsgálatokat. Ezen második szakasz kezelése a normal.ini, illetve ahhoz hasonló nevű fájllal történik. Bár erősen technikai jellegű, ezen konvenciók ismerete nélkülözhetetlen ahhoz, hogy a vonatkozó fejlesztéseket megérthessük.

¹⁶ https://nimbus.elte.hu/bbgc



6. ábra: A Biome-BGCMuSo fájlrendszere. ini: a modellbeállításokat tartalmazó fájl. restart: a modell egyensúlyi beállításait tartalmazó fájl (az ún. spinup végeredménye). met: meteorológiai input adatokat tartalmazó fájl. CO2: légköri CO2 koncentrációt leíró fájl. mgm: az emberi beavatkozást leíró vezérlő fájl. soil: talajfájl. epc¹⁷: ökofiziológiai paramétereket tartalmazó fájl. Ndep: légköri nitrogén ülepedést leíró fájl. planting¹⁸, thinning, mowing, grazing, harvesting, ploughing, fertilizing, irrigation, mulching, cwd-extract: a megfelelő emberi beavatkozást leíró fájlok (rendre vetés, ritkítás, kaszálás, legeltetés, aratás, szántás, trágyázás, mulcsolás, holtfa eltávolítás).

Tekintve az ökofiziológiai paraméterek fontosságát a Mellékletben bemutatok egy példát a fájl natív formátumában (M1. táblázat). Amikor a paraméterek beállításáról, vagy más szóval inverz paraméter-becslésről írok a Biome-BGCMuSo kontextusában, akkor valójában ennek az EPC fájlnak adott sorait szeretném beállítani. Az Irodalmi áttekintés fejezetben említett "dimenzionalitás átka" itt nyilvánvaló, hiszen, ha több paramétert szeretnénk egyszerre beállítani, az önmagában óriási módszertani feladat.

¹⁷ Két helyen is megtalálható ez a fájlhivatkozás. Az INI-ben a vetés előtti növényt, a "planting"-ben pedig a vetést követő növényt jellemzi.

¹⁸ A modellben a "planting" fájl nem csak az ültetést, hanem a vetést is jelentheti. A modell általánosságának következménye, hogy a fájlnevek nem lehetnek teljesen pontosak.

A modell számos kimenő változót származtat a futás során. Ezen kimenő adatok számkódokkal vannak definiálva, és egy konverziós tábla segítségével tudja a felhasználó kiválasztani az általa fontosnak tartott eredmények listáját. Például az egy oldalas levélfelület index (LAI) kódja 2520, míg a napi szintű GPP kódja 3009.

3.2. A Biome-BGCMuSo köré épülő keretrendszer: RBBGCMuso

A Biome-BGCMuSo modell az eredeti Biome-BGC modell alapfilozófiáját örökölte, ami azt jelenti, hogy kizárólag szöveg alapú beállító fájlok segítségével, parancssoros módban használható. Ennek közvetlen következménye, hogy eltekintve néhány triviális esettől, programozói tudás hiányában előnyeinek kiaknázásához a szimulációhoz kapcsolódó fájlok manuális elő- és utófeldolgozása szükséges. Ezek idő- és energiaigényes természete korlátozza a modell tudományos vagy döntéstámogatói célú alkalmazását. Például, ha a vizsgálat tárgya egy tetszőleges paraméter változtatásának hatása az előrejelzett kukoricatermésre, kétszer szükséges futtatnunk a modellt, két manuális változtatást kell alkalmaznunk a bemeneti fájlokon, majd kétszer kell beolvasnunk az eredményeket, majd konvertálnunk és ábrázolnunk kell azokat.

Ezzel szemben az ehhez hasonló vizsgálatok programozói tudás birtokában nagymértékben automatizálhatók, ezzel spórolva időt, és így téve lehetővé olyan problémák megoldását és kényelmi funkciók bevezetését is, amelyek korábban megoldhatatlannak bizonyultak (Bagnara et al., 2019). Ezáltal lehetőségünk nyílik például arra, hogy kiválasszuk a modellezés kontextusában meghatározó ökofiziológiai paramétereket. Ugyancsak lehetővé válik minden fontos változó hatását egyidejűleg vizsgáljuk a célváltozóra egy előre meghatározott intervallumban, ezzel téve lehetővé olyan összetett összefüggések felfedezését is, amelyek a paraméterek önálló, egyesével történő tanulmányozásával nem érzékelhetők (pl. érzékenység-vizsgálat, modell-optimalizáció, vagyis a modell-adat fúzió fontosabb elemei; lásd. a 2.3. fejezetet).

A modellhez programok, szkriptek, illetve eljárások írása azonban bonyolult folyamat, amelynek a végén ezek futtatását gyakran kizárólag szerzőjük képes elvégezni. Így ezek a megoldások sok esetben nem reprodukálhatóak. Erre a problémára kínálnak megoldást olyan szoftvercsomagok, amelyek a gyakran ismétlődő mintázatokat általánosítják, ezzel téve egyszersmind egyszerűbbé, koherensebbé, valamint konzisztensebbé a programozási feladatokat.

Doktori munkám megkezdése előtt a Biome-BGCMuSo modellhez programozási keretrendszer nem állt rendelkezésre, leszámítva azt a szintén saját fejlesztésű R-csomagot (Hollós et al. 2023), amely akkoriban még kezdetleges állapotban volt, s amelynek teljessé tétele doktori munkám egy jelentős eredménye. A szoftvercsomag fejlesztésénél fontos szempont volt olyan programozási nyelv kiválasztása, amely tükrözi a potenciális modellhasználók preferenciáit. Itt

28

lényegében két lehetőségem volt: az R (R Core Team, 2021), vagy a Python programozási nyelvek használata. Ezek közül a statisztikai célú R-t választottam, mivel a fejlesztések megkezdésekor jóval gazdagabb adatelemzői és statisztikai eszköztár jellemezte ezt.

A szóban forgó R csomag fejlesztésének legfőbb szempontjai az átláthatóság, valamint a jövőbeli kooperatív fejlesztések lehetőségének megteremtése voltak. Ezt korszerű verziókövető rendszer segítségével valósítottam meg (Git), amihez a szervert a GitHub szolgáltatta (<u>https://github.com/hollorol/RBBGCMuso</u>). Fontos megemlíteni, hogy ma már elvárás a különböző szoftvercsomagok közzététele a GitHub-on. Giten keresztül a különböző verziójú kódok fejlesztését egyidejűleg lehet végezni, ezzel garantálva a kódminőséget, a változások követhetőségét, valamint azt, hogy több fejlesztő esetén ne legyenek ütközések a különböző fejlesztőknél lévő kódok között. A GitHub továbbá lehetőséget nyújt nemzetközi kommunikációra és kód-népszerűsítésre is.

3.2.1. Az RBBGCMuso felépítése

Az RBBGCMuso csomag az úgynevezett funkcionális programozási paradigma elvei szerint épült fel (Khanfor & Yang, 2017). Ez az elv az erre képes nyelvek esetén garantálja, hogy minden egyes függvény "tiszta" legyen, vagyis adott bemeneti adat egyértelműen determinálja minden függvény esetén a kimeneti adatot. A hozzárendelés ez esetben egyértelmű, tehát minden függvény a funkcionális nyelvekben, matematikai értelemben is függvény. Ezzel elkerülhetjük az olyan eseteket, amikor ugyanolyan bemeneti feltételek mellett különböző végeredményeket kapunk, ezzel biztonságossá válik a függvények használata. Ezzel a csomaggal válik igazán reprodukálhatóvá bármilyen modellfuttatás.

A modellt elsődlegesen az INI fájlok vezérlik (Hidy, 2021). Ebből a fájlból kiindulva minden lényeges információ megtalálható a modell beállításairól, a felhasznált fájlokról stb. (6. ábra). Az RBBGCMuso alapvetően ezen szöveges fájlok kezelésével foglalkozik, majd a modell futtatásával, a futtatás eredményeinek beolvasásával adatstruktúra formájában, és az eredmények értelmezésével hatékony módon. A Biome-BGCMuSo-ban a változók (paraméterek, beállítások) adatokkal történő összekötése annak pozíciójával történik a fájlokban (azaz annak leírásával, hogy hányadik sorban szerepel). Mivel ez a pozíció a gyakori fejlesztések során többször változott, az RBBGCMuso-ban egy verziófüggő sorszám a kapcsolódási pont. Ez alól kivételt jelentenek az allokációs paraméterek és egyéb vektor-típusú változók, mivel ezen esetekben lehetséges, hogy a változó a sor közepén található (mint például a virágzás fenofázisának hossza, aminek az értéke 75 hőfok-nap a látható; M1 táblázat). A változók indexelésére egy olyan módszert találtam ki, amely kompatibilis a régi verziókkal és az újabbakkal is, valamint nem igényel karekterlánc-műveleteket (vagyis gyors és hatékony). Amennyiben adott változó (v) az EPC fájl r. sor c. oszlopában van, az indexe a következő összefüggéssel írható le: $i_v = r + (n - 1)/10 + (c - 1)/100$, ahol n az összes oszlop száma. Például a 7. ábrán a virágzás hőfok-napban kifejezett változójának indexe 128.64, mivel a 128. sorban található, 7 fenofázis van, és azok közül az ötödik vonatkozik a virágzásra. Az általam bevezetett módszer előnye, hogy egyszerű aritmetikai műveletekkel elvégezhető a beazonosítás, továbbá az olyan változók esetén, mint például a C:N arány a termésben, ahol csak 1 elem található a sorban a szokásos sorszám ugyanúgy alkalmazható (29.00 = 29).

PHENOLOGI	CAL (ALI	LOCATION) PARAME	TERS (7 phe	enological	phases)		
csirazas	keles	<pre>levelnov_felfut</pre>	levelnov	viragzas	szemtelitodes	erett	(text) name of the phenophase
10	50	240	240	75	850	10000	(Celsius) length of phenophase (GDD)

7. ábra. Több változó kezelése egy sorban az EPC fájlban.

Az RBBGCMuso részére a felhasználó kell biztosítsa a modell végrehajtható állományát, illetve a működőképes formában összeállított bemenő adatfájlokat (INI, EPC, meteorológia, MGM stb.). A szoftver elvárása, hogy minden fájl egy adott mappába kell legyen szervezve, alkönyvtárak nélkül. Az RBBGCMuso részeként megírt *flatMuso* függvény ezt a felhasználó számára meg tudja oldani, amennyiben szükséges (ami egy kényelmi szolgáltatás).

A fájlrendszerre vonatkozóan az egyszerűség kedvéért saját szabályokat definiáltam, amelyek megkönnyítik a függvények használatát, leegyszerűsítik a beállításokat így érve el sok esetben magas szintű funkciókat mindössze egyetlen sor R-kóddal. A konvenciók a következők:

- Minden szükséges fájl ugyanabban a mappában, ugyanazon a szinten található meg a fájlrendszerben.
- A normal.ini fájlok neve n.ini-re végződik (például martonvasar_n.ini), a spin up ini fájlok neve s.ini-re végződik (martonvasar_s.ini).
- A normal.ini fájlban a kimeneti fájl bináris típusra van állítva.

A konvenciók betartásával a modell az adott mappában már közvetlenül futtatható, ábrázolható, kalibrálható, így további beállítást nem igényel.

A modellcsomag előkészítése után kap az RBBGCMuso központi eleme szerepet. A szoftver logikája szerint az INI fájl-ból a vezérlő információkat (fájlok helyei a számítógépen, futás időszaka, kimeneti változók listája stb.) a setupMuso függvénnyel lehet kinyerni, aminek eredménye egy strukturált lista (8. ábra).

A modell beállító struktúrájára építenek a modellt spinup (*spinupMuso*), illetve normal módban (*normalMuso*) futtató központi függvények. Amennyiben nem csak futtatni szeretnénk a modellt, vagy szeretnénk a két módot egymás után végrehajtani, a *runMuso* kényelmi függvény áll rendelkezésünkre (a legfontosabb függvények kompozícióját a 9. ábra mutatja be). Ez úgy tudja

futtatni a modellt tetszőleges módban, hogy annak egy kiválasztott fájljában módosítást eszközöl (a *changeMuso* függvény által).

😊 settings	list [19]	List of length 19
executable	character [1]	'/home/bzoli/CIRM_cikk/muso'
calibrationPar	double [1]	1
outputLoc	character [1]	'/home/bzoli/CIRM_cikk/.'
outputNames	character [2]	'site' 'site'
inputLoc	character [1]	'/home/bzoli/CIRM_cikk'
iniInput	character [2]	'n.ini' 'n.ini'
🔘 metinput	character [2]	'Martonvasar.wth' 'Martonvasar.wth'
epcInput	character [2]	'maize.epc' 'maize.epc'
inputFiles	character [2]	'n.ini' 'n.ini'
numData	double [3]	265720 8736 728
startYear	double [1]	1991
numYears	double [1]	28
outputVars	list [2]	List of length 2
soilFile	character [2]	'Martonvasar.soi' 'Martonvasar.soi'
dailyVarCodes	character [26]	'2520' '3009' '3054' '3103' '2794' '3002'
annualVarCodes	character [7]	'3050' '3022' '3029' '3061' '2734' '3116'
dailyOutputTable	list [26 x 3] (S3: data.frame)	A data.frame with 26 rows and 3 columns
annualOutputTable	list [7 x 3] (S3: data.frame)	A data.frame with 7 rows and 3 columns
normOutputFlags	double [2]	10

8. ábra. A setupMuso parancs eredménye R-ben: példa a modellfuttatást leíró központi adatstruktúrára.



9. ábra. Biome-BGCMuSo futtatására, a paraméterek megváltoztatására használt függvények kompozíciós sémája.

A csomag meghatározó eleme a párhuzamos futtatást lehetővé tevő mechanizmus, ami az ún. future¹⁹ R csomagra épül (Bengtsson, 2021), és az aszinkron programozás alapelvét használja. A párhuzamosítás kapcsán fontos megemlíteni, hogy ún. lock fájl alapú mutex segítségével szinkronizál, és megoldottam a folyamatjelző sáv (angolul "progress bar") megjelenítését is, amivel

¹⁹ https://cran.r-project.org/web/packages/future/index.html

a felhasználó információt kap a futás állapotáról (ne felejtsük el, hogy itt adott esetben több 100 000 vagy milliós futásszámról van szó, tekintve a probléma nagy dimenzionalitását). A párhuzamos futtatás megvalósítására korábban nem létezett megoldás a Biome-BGC kontextusában. A több szálon, illetve processzoron való futtatás révén nagyságrendekkel rövidebb idő alatt fut le egy nagyszámú szimuláció.

Ezen alapvető függvényekre épülnek a további, magasabb szintű használatot támogató lehetőségek.

Funkcionalitás szempontjából az RBBGCMuso csomag függvényei 8 csoportba sorolhatók (terjedelmi okokból itt csak a függvények nevét sorolom fel):

- 1. Futtatás környezetét definiáló, beállító függvény (setupMuso)
- Modellfuttatást végző függvények (spinupMuso, normalMuso, rungetMuso, runMuso)
- Elő- és utófeldolgozó függvények (setupMuso, changemulline, flatMuso, alignData, cleanupMuso, corrigMuso, fixAlloc, getyearlycum, getyearlymax, postProcMuso, putOutVars)
- 4. Ellenőrzést végző függvények (checkFileSystem, checkMeteoBGC)
- 5. Ábrázoló függvények (compareMuso, plotMuso, plotMusoWithData)
- Adatlekérő, inputfile készítő függvények (getAnnualOutputList, getConstMatrix, getDailyOutputList, getFilePath, createSoilFile, genEpc, getFilesFromIni, getLogs, getOutFiles, getSoilDataFull, readObservedData, saveAllMusoPlots)
- Tudományos funkcionalitást biztosító függvények a modell-adat fúzió kontextusában (calibrateMuso, compareCalibratedWithOriginal, multiSiteCalib, musoGlue, musoMonte, musoQuickEffect, musoRand, musoSensi, optiMuso, paramSweep, tuneMuso).
- Segédfüggvények (copyMusoExampleTo, dynRound, fextension, multiSiteThread, musoCompareFiles, musoDate, musoGetValues, numcut, numcutall, postProcMuso, readErrors, readValuesFromFile, stampAndDir, stampMuso, tuneMusoServer, tuneMusoUI)

Az Eredmények fejezetben a fontosabb, gyakorlati szempontból kiemelt jelentőségű funkciókat mutatom be. Az alábbi alfejezetben a Monte Carlo alapú randomizáció algoritmusát részletezem, ami újdonságtartalma miatt kiemelendő, és máshol nem került dokumentálásra.

3.2.2. Az RBBGCMuso randomizációs algoritmusa

Mivel az RBBCMuso egyik legfontosabb része a Monte Carlo alapú, véletlenszámok generálásán alapuló szimuláció-sokaság végrehajtása, ezért kiemelt figyelmet fordítottam ennek az RBBGCMuso-ban való precíz megvalósítására. A probléma összetettségét jellemzi, hogy bizonyos modellparaméterek nem függetlenek egymástól (lásd 3.1. fejezet), és ez a randomizáció során egy megoldandó komoly matematikai feladat.

A modellben megjelenő paraméter-összefüggések leírhatók a paraméterek ($\theta = (\theta_1, ..., \theta_{n_p})$) lineáris kombinációira vonatkozó feltételek segítségével (a Biome-BGCMuSo modell kontextusában ezek a paraméterek általában a korábban leírt ökofiziológiai paraméterek). Ez általánosan azt jelenti, hogy a paramétertér lineáris transzformációira vonatkozóan feltételeket kell szabnunk. Ennek két fő komponense az egyenlőségi feltételre vonatkozó transzformáció (*E* mátrix és a hozzá tartozó \overline{e} feltétel), valamint az egyenlőtlenségi feltételre vonatkozó transzformáció (*G* mátrix és a hozzá tartozó \overline{h} feltétel).

A fentebb leírt konvex teret (azaz politópot) az általam választott megoldás során a paramétertérből történő egyenletes mintavételezéshez a Hit & Run algoritmust (Zabinsky & Smith, 2013) alkalmaztam. A Biome-BGCMuSo paraméter-függőségei konvex típusúak, ami azt jelenti, hogy leírhatóak a következő egyenletek/egyenlőtlenségek segítségével:

$$\mathbf{G}\mathbf{\theta} \le \overline{h}$$
 (2)

$$\mathbf{E}\boldsymbol{\theta} = \overline{\boldsymbol{e}} \tag{3}$$

A módszer 3 fő lépésből áll:

- 1. Egy kezdeti paraméterhalmaz kiválasztása, amely kielégíti a feltételeket
- 2. Véletlen irány választása egy a ponton (egy paraméterhalmaz egy pontnak felel meg) áthaladó egyenes meghatározásához
- 3. Mivel a politóp konvex, az egyenes összeköt két olyan határpontot, amely között minden pontja az egyenesnek a politópon belülre esik. Ezen lehatárolt szakaszban véletlenül kiválasztott pont lesz a következő minta.

Az algoritmus konvergenciáját Gelman-Rubin teszttel (Gelman et al., 1995) lehet tesztelni. A (3.) egyenlőségi feltételeket tartalmazó **E** mátrix konstruálását a leggyakoribb függőségi példán keresztül fogom bemutatni: amikor a paraméterek összegének egynek kell lennie. Erre jó példa az allokációs paraméterek. Ezek olyan arányszámok, amelyek a modellben megmutatják, hogy egy adott fenofázisban az allokált anyagmennyiség hányad része jut egy adott növényi részbe (például szárba; lásd 3.1. fejezet). A probléma amennyiben mindegyik paramétert a [0,1] intervallumról mintázzuk, egyszerű n-dimenziós Dirrichlet-mintázás (tekintve, hogy a terünk egy n-dimenziós szimplex).

Hasznos azonban a mintavételezési problémát konvex-politóp mintázásnak tekintenünk, mivel a paramétereinkre vonatkoznak különálló értékmegkötések. Ehhez a mátrix előállításához tekintsük az alábbi egyszerűbb esetet három paraméterre vonatkozóan: p_1 , p_2 and p_3 :

$$p_1 + p_2 + p_3 = 1 \tag{4}$$

Az egyenlőségi példában (a 3. egyenletben) az *E* mátrix az (111), az \overline{e} vektor pedig az egy elemű (1) vektor:

$$(1 \quad 1 \quad 1) \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = (1)$$

Ezzel egyidőben ezen paraméterek mindegyikének egy előre meghatározott intervallumban $([a_i, b_i])$ kell lennie.

$$a_{1} \leq p_{1} \leq b_{1}$$

$$a_{2} \leq p_{2} \leq b_{2}$$

$$a_{3} \leq p_{3} \leq b_{3}$$
(5)

A fenti 5. egyenlőtlenségek a következő ($G\theta \leq \overline{h}$ formájú) mátrix egyenlőtlenséggel (2. egyenlet) is reprezentálhatóak:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \leq \begin{pmatrix} b_1 \\ -a_1 \\ b_2 \\ -a_1 \\ b_3 \\ -a_1 \end{pmatrix}$$
(6)

Általánosan a paraméter intervallumainkból származó megkötések reprezentálhatóak a következő formában:

$$[I; -I]\theta \le \left[\overline{u}; -\overline{l}\right] \tag{7}^{20}$$

²⁰ Az [A; B] jelölés a mátrixok, vagy oszlopvektorok vertikális összefűzését jelöli itt, míg az "I" egységesen az egységmátrixot

Egy további gyakori megkötés a Biome-BGCMuSo esetében, mikor a paraméter értékek a " \leq " rendezési relációban állnak egymással. Erre példa a C:N arány a különböző növényi részekben, amelyekre vonatkozóan az a megkötés, hogy a levelekben a legalacsonyabb ez az érték. A fenti általános példát alapul véve három paraméterre ($p_1 \leq p_2 \leq p_3$), ahol a következő relációknak kell teljesülnie, a 2. egyenlet a jelen példába behelyettesítve a következő:

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \leq \begin{pmatrix} b_1 \\ -a_1 \\ b_1 \\ -a_1 \\ b_1 \\ -a_1 \\ 0 \\ 0 \end{pmatrix}$$

Általánosabban, ha $p_i \le p_j (i \ne j)$, akkor van olyan sor a **G** mátrixban, amelynek i. sorában 1, j. sorában -1 áll, mindenhol máshol pedig 0. Ha ezen sor indexe k, akkor a jobboldali oszlopvektor k. eleme 0.

A fentebbi példa szerinti szükséges átalakításokat az RBBGCMuso csomag automatikusan előkészíti az limSolve csomag (Meersche et al., 2009) Hit and Run samplinget megvalósító függvényei számára. A modellre vonatkozó megkötéseket JSON-beállítófájlban tárolom, ami egy egyszerű és széles körben használt szöveges állomány különböző adatstruktúrák leírására. A modell verzióváltásai általában új beállítófájlokat tesznek szükségessé. Az RBBGCMuso csomagban a 4. Biome-BGCMuSo verziótól egészen a 7. verzióig megtalálhatóak a beállítófájlok, így lehetőségünk van igény szerint régebbi modellverzióval is dolgozni.

A paraméterekre vonatkozó intervallum-megkötéseket az RBBGCMuso csomag a parameters.csv fájlból veszi, amiben a paraméter neve, indexe (ld. 30. oldal), valamint minimum és maximum értékei találhatóak.

3.3. Megfigyelési adatok

Munkám során a kukorica terméshozamának állomás szintű²¹ megfigyelései Martonvásárra (47°19'56,97"N, 18°47'50,61"E) vonatkoznak. Martonvásár éghajlata kontinentális, mediterrán és óceáni hatásokkal. Az évi középhőmérséklet 11,2 °C, a hosszú távú átlagos csapadékmennyiség 550 mm körül van. A FAO-WRB osztályozási rendszer szerint (IUSS munkacsoport, 2015) a terület talajtípusa a Haplic Chernozem, átlagosan 51,4% homok-, 34% iszap- és 14,6% agyagtartalommal.

²¹ Pontbeli, nem térben aggregált adatforrás, szemben a később említésre kerülő NUTS3-szintű adatsorokkal.

A térfogattömeg 1,47 g cm⁻³, a pH 7,3, a CaCO₃ tartalom 0-1%, a talaj szervesanyag-tartalma a feltalajban átlagosan 3,2%. A talaj növény számára hozzáférhető makrotápanyag-ellátottsága a ProPlanta növénytáplálási tanácsadó rendszer alapján a foszfor tekintetében gyenge, a kálium tekintetében közepes vagy jó (Fodor et al., 2011). A mezőgazdasági tartamkísérletekben minden kezelés véletlenszerű blokktervben van elrendezve, 20-40 m-es² parcellákkal, négy ismétlésben.

A kísérleti adatokat a Mezőgazdasági Kutatóközpontban az 1960-as években létrehozott hosszú távú szántóföldi kísérletek (LTFE) során gyűjtötték. A kukoricát Martonvásár városában és környékén több LTFE helyszínen termesztik. A kísérletek egy része a szerves és ásványi műtrágya kijuttatásának a terméshozamra gyakorolt hatását vizsgálja. Mások a talajművelés, a hibridválasztás, a vetésidő és a növénysűrűség hatását vizsgálják a termés élettanára és a terméshozamra. Jelen tanulmányban egy összetett kukoricatermés-adathalmazt használtunk fel, amely a FAO350-400 hibridek termésadatainak lekérdezésével készült az 1994-2018 közötti időszakban a magyar kukoricatermesztési rendszerre jellemző négy LTFE magas nitrogénszintű (legalább 160 kgN/ha⁻¹év⁻¹) kezelésében (növénysűrűség: 70.000 növény/ha; vetés időpontja: április második dekádja; betakarítás időpontja: október közepe). A termésátlagot a négy kiválasztott LTFE 16 termésadatából számoltuk ki (minden LFTE-ben 4 ismétléssel, 20-40 m-es parcellamérettel) minden évben. A kukorica átlagtermése (az összes parcellából és az összes évből kiszámítva) a kísérletekben 7,7 t ha⁻¹ volt, viszonylag nagy bizonytalansággal (SD = 1,58 t ha⁻¹; min = 0,96 t ha⁻¹; max = 13,57 t ha⁻¹). Megjegyzendő, hogy a megfigyelések a szemtermés szárazanyagára vonatkoznak.

A tanulmányban a kísérleti helyszínt magába foglaló Fejér megyei kukoricatermés-adatokat is felhasználtam. A megfigyelési adatokat a Központi Statisztikai Hivatal adatbázisából nyertük az 1991-2018 közötti időszakra vonatkozóan. A kukoricatermés átlaga a vizsgált időszakban Fejér megyében 6,36 t ha⁻¹ volt (SD = 1,92 t ha⁻¹; min = 2,89 t ha⁻¹; max = 10,06 t ha⁻¹).

3.4. Klímaadatok

A FORESEE az ELTE Meteorológiai Tanszék által létrehozott és fenntartott meteorológiai adatbázis, amelyik egyedi módon kombinálja a múltra vonatkozó meteorológiai alapadatokat, és a jövőre vonatkozó projekciókat. A FORESEE létrehozását az éghajlatváltozással összefüggő hatásvizsgálatok kapcsán felmerülő meteorológiai adatigény motiválta, amelyre a korábbi években nem volt hazai megoldás. A FORESEE érdekessége, hogy a származtatott rácsponti adatai közvetlenül felhasználhatók a dolgozatban bemutatott és a munkám során felhasznált Biome-BGCMuSo modell bemenő adataként. A FORESEE szabadon hozzáférhető bárki számára.

A FORESEE interpolált napi minimum- és maximum-hőmérsékletet, illetve napi csapadékösszeg adatokat tartalmaz a Kárpát-medence tágabb térségére, az 1951–2100-as időszakra
vonatkozóan. Az alapadatokat állomásszintű megfigyelési, és hiba-korrigált (angolul bias-corrected) modellezett adatok adják. A FORESEE adatbázis első verziója 2015-ben került kiadásra 1/6° × 1/6°- os térbeli felbontással, és az ENSEMBLES FP6 projekt A1B szcenárióra vonatkozó 10 modell eredményeinek felhasználásával (Dobor et al., 2015). Ennek a gyűjteménynek az utolsó verziója a 3.2-es volt. 2022-ben két új verziót hoztunk létre, melyek immár egy javított, a CarpatClim és a HUCLIM adatbázisokkal megegyező rácson, 0,1° × 0,1°-os térbeli felbontással tartalmaznak adatokat egy tágabb területre vonatkozóan (FORESEE v4.0; Kern et al., 2024). Az adatbázisra vonatkozó visszajelzések alapján készült egy csak Magyarország területére vonatkozó, módosított adatsort is (ennek a neve FORESEE-HUN, és jelenleg az 1-es verziószámmal láttuk el). A FORESEE v4.0, illetve FORESEE-HUN v1.0 jövőbeli éghajlatát az EURO-CORDEX adatbázis 14 regionális éghajlati modellje vetíti előre az RCP4.5 és RCP8.5 forgatókönyveinek felhasználásával. Az eredeti elképzelések logikáját megtartva a FORESEE alkalmas az éghajlatváltozás mértékének és várható hatásainak *ensemble* alapú megközelítésre is. A FORESEE legutóbbi verziójának az elkészítésében és frissítésében én is részt vettem (Kern et al., 2024).

A fent említett alapvető klímaadatokon túl a FORESEE tartalmaz nappali átlagos vízgőz telítési hiányt (VPD), nappali átlaghőmérsékletet, és nappali átlagos rövidhullámú sugárzás adatokat is, melyeket az MT-CLIM modell segítségével származtattunk (Hungerford et al., 1989).

Mivel a dolgozatban bemutatott munka évekkel ezelőtt kezdődött, értelemszerűen nem volt módom a jelenleg elérhető, legfrissebb FORESEE adatbázis használatára. Emiatt a modelloptimalizácó során a FORESEE 4.0 verziót, míg az országos, kukoricára vonatkozó futtatások során a 3.2 verziót használtam.

3.5. Talajadatok

A DOSoReMI (Digital, Optimized, Soil Related Maps and Information in Hungary) adatbázist (Pásztor et al., 2020) az az igény hozta létre, hogy megpróbáljon hazai és nemzetközi igényeket optimálisan kielégítő, országos léptékű, digitális, tematikus talajtérképeket szolgáltatni a kapcsolódó tudományterületek kutatásaihoz. Az alkotók különböző talajtani felvételezések, térképezések során gyűjtött aktuális és archív adatok, hagyományos talajtérképek és a talajok szempontjából meghatározó környezeti változók tematikus és térbeli kapcsolatait elemezték. A feltárt összefüggéseket a digitális térképezés számára formalizálták és térinformatikai és geostatisztikai eszközök segítségével elvégezték az egyes, a talajok állapotára, folyamataira és funkcióira jellemző talajtani változók térbeli kiterjesztését. Az eredmény egy 100×100 méteres térbeli felbontású adatrács, amely minden cellára tartalmazza a talaj 0-30, 30-60, 60-100, 100-150 és 150-200 cm mélységben lévő rétegeire (többek között) az alábbi adatokat: térfogattömeg (g/cm³), szervesanyag-tartalom (%), mésztartalom (%), pH, homok/iszap/agyagfrakció (%), maximális vízkapacitás

(m³/m³), szabadföldi vízkapacitás (m³/m³), hervadásponti nedvességtartalom (m³/m³), termőréteg vastagsága (cm).

3.6. Paraméter optimalizáció

A modell-adat fúziónak (2.3 fejezet) a legfontosabb, és egyben legnehezebb része az inverzmodellezés. Éppen ezért disszertációm egyik központi eleme a paraméter-optimalizáció. A terület 2.3.2. fejezetben említett problémáira dolgoztam ki egy új modell-kalibrációs eszközt, amit feltételes tartománycsökkentő módszernek²² neveztem el. Ebben a fejezetben ennek fogom bemutatni a módszertani alapjait, valamint az elkészült algoritmust is fel fogom vázolni.

3.6.1. A feltételes tartománycsökkentő módszer

3.6.1.1. Az algoritmus leírásának elméleti háttere és jelölésrendszere

Az inverz modellezés irodalmi áttekintése során már érintettem az elméleti hátteret, azonban a feltételes tartománycsökkentő módszer bevezetéséhez és a szükséges jelölésrendszer bemutatásához szükség van a témakör részletesebb bemutatására és a szükséges matematikai formalizmus bevezetésére.

Legyen $\mathcal{M}: \mathcal{S} \mapsto \mathcal{O}$ egy tetszőleges folyamatorientált modell, amelynek kimenete \mathcal{O} , lehetséges parametrizációinak halmaza pedig \mathcal{S} ($\theta \in \mathcal{S}$, ahol $\mathcal{S} \subset \mathbb{R}^n$ az n dimenziós bemeneti paramétertér, és $\mathcal{O} \subset \mathbb{R}^{l \times k}$ a modell kimeneti mátrixa, ahol l *az* időlépések száma és k *a* kimeneti adatfolyamok száma (amelyek a szimulált változókat képviselik)). Megjegyzendő, hogy ebben a tanulmányban a "paraméter" szó állítható bemeneti adatokat jelent. Egyes esetekben konvex kényszereket kell definiálni az \mathcal{S} halmazon (pl. a bemeneti paraméterek közötti függőségek kezelésére; lásd. 3.2.2 fejezet, amelynek idevonatkozó jelölésrendszerét alkalmazzuk).

Amennyiben a mérési adatok (*D*) és bizonytalanságaik (az adatokra vonatkozó sűrűségfüggvény) elérhetőek, valamint ismert a modellezett rendszer és a paramétereinek valószínűségi összefüggése (likelihood), a modellező közvetlenül tud a bemeneti paraméterekhez kapcsolódó sűrűségfüggvényből (pdf, $p(\theta|d), d \in D$) mintázni²³. Ez a pdf az úgynevezett *a posteriori* sűrűségfüggvény (innentől röviden poszterior függvény), ami intuitív módon, ha a paramétertér egy folytonos véletlen változóval írható le, akkor kifejezi annak a valószínűségét, hogy

²² Angolul Condition Interval Reduction Method (CIRM)

²³ p-vel jelölöm a sűrűség- és a súlyfüggvényt is. A legáltalánosabb mértékelméleti megközelítést alkalmazom: együtt kezelem a folytonos/diszkrét/kevert eseteket.

 θ a (θ , θ + ε) intervallumba esik a mért adatok ismeretében (d), ahol ε egy végtelenül kicsi vektor. A poszterior függvény segítségével az inverz modellezés minden célja kielégíthető.

Csakhogy általában a poszterior függvény nem ismert (a korábbi tapasztalatok érvénytelenek lehetnek az új optimalizálási feladatokra; Van Oijen et al., 2005) és explicit nem leírható. Ezzel szemben a likelihood függvényre ($L = p(d|\theta), d \in D$; azaz a modellhiba valószínűségi modelljére; Gelman et al., 1995) vonatkozóan léteznek általános ajánlások. Ilyen például a 2.3.2. fejezetben ismertetett normál eloszlású likelihood függvény használata (ennek korlátairól később írok). A poszterior függvény és a likelihood függvény közti kapcsolatot a Bayes-tétel biztosítja (8. egyenlet).

$$p(\theta|d) = \frac{p(\theta|d)p(\theta)}{p(d)}$$
(8)

A tétel megmutatja, hogy a paraméterekre vonatkozó előzetes tudásunkat ($p(\theta)$ (az úgynevezett *a priori* sűrűségfüggvény; innentől röviden prior²⁴) felhasználva módosítsuk az adott adatra vonatkozó likelihood értéket, hogy így kapjuk meg a paraméterhez tartozó poszterior valószínűséget. A tételben a számlálóban szereplő adatokra vonatkozó valószínűség nem függ a paraméterektől és a Monte Carlo mintavételezés minden egyes lépésében állandó (9. egyenlet).

$$p(d) = \int p(\theta)p(\theta|d)d\theta = c \in \mathbb{R}$$
(9)

Szerencsére, a legtöbb Monte Carlo módszer (reject sampling, MCMC algoritmusok) csupán a $p(\theta|d)p(d)$ értékek ismeretében tudnak a poszteriorból mintázni, így nem kell meghatároznunk az amúgy nehezen meghatározható p(d)-t.

A Bayes-féle munkafolyamat kritikus része a paramétereinkre vonatkozó előzetes tudásunk felhasználása a megfelelő prior függvény kiválasztására. A leggyakoribb előzetes tudás "a tudás hiányának" tudása. Ezekben az esetekben a priorok nem adnak többlet információt a valószínűségi modellezéshez. Az ilyen priorokat nem-informatív priornak hívjuk, s leggyakrabban alkalmazott ilyen eloszlás az egyenletes eloszlás ($p(\theta_i) \sim U(I_{i,1}, I_{i,2})$). Az egyenletes prior viszont nem mindig megfelelő, mert nem invariáns a paraméter léptékére, valamint túlságosan érzékeny a meghatározó alsó ($I_{i,1}$) és felső korlátokra ($I_{i,2}$).

Gyakorlati szempontból a megfelelő prior kiválasztása segíthet elkerülni a túlillesztést. Az egyenletes prior használata azonban nem változtatja meg a ML értékek helyét, így ezek használatával az ML becslésnél tapasztalt túlillesztést sem kerüljük el (Gelman, 1996). Nagy mennyiségű megfigyelési adat esetén a prior kiválasztása kevésbé fontos, és a maximum-likelihood becslések (megközelítőleg) megegyeznek a maximális poszterior becslésekkel, függetlenül a priorfüggvény kiválasztásától.

²⁴ ld. 2.3.2. fejezet

A nyilvánvaló hátrányok ellenére az egyenletes priorokat gyakran használják a Bayes módszerek esetén, mivel egyszerűek, és mert könnyebb a paraméterekről egy korlátos térben gondolkodni. (Gelman, 1996; Stedinger et al., 2008; Gelman és Yao, 2020; Wallach et al., 2021).

A továbbiakban egy olyan új módszert mutatok be, amely a mind a Bayes, mind a frekventista modelloptimalizációs módszert kiterjeszti a gyengeségeik kezelésének céljából. Dolgozatomban a GLUE módszert használtam demonstrációs céllal és az egyszerűbb interpretálhatóságot figyelembe véve (kihasználva a GLUE-ra jellemző vizualizációs módszert is (lásd. 4. ábra), ahol az ekvifinalitás problémája a marginális eloszlások esetén bemutatható). A következő részben az új módszer részletes leírása következik.

3.6.1.2. Szűrőfüggvények segítségével történő tartománycsökkentés

Legyen $f: \mathcal{O} \to \{0, 1\}$ egy olyan függvény, ahol $f \circ \mathcal{M}^{25}$ a modell paramétereket két osztályba sorolja: elfogadható (1) és nem elfogadható (0) kimenet. Az elfogadható ebben az összefüggésben azt jelenti, hogy \mathcal{O} összhangban van a modellező elvárásaival. Szeretnénk hangsúlyozni, hogy az elfogadhatóság megítélése nem közvetlenül a megfigyelés és a szimuláció mennyiségi összehasonlítása alapján történik, hanem a szimulált rendszerről való néhány további ismeret alapján (például NPP/GPP arány, vagy maximális gyökerezési mélység). Ez a fajta tudás származhat a tudományos irodalomból, vagy a modellező mindennapi gyakorlatából. A modellt úgy kell beállítani, hogy \mathcal{O} tartalmazza a folyamatra vonatkozó információkat, amelyekre vonatkozóan fszűrő-függvényként viselkedik. Formalizálva:

$$\theta \in S \text{ elfogadhat} \acute{} \Leftrightarrow (f \circ \mathcal{M})(\theta) = 1 \tag{10}$$

Mivel csak azok a paraméterértékek képzik érdeklődésünk tárgyát, amelyek elfogadható kimenetet biztosítanak (a továbbiakban ezeket elfogadható paramétereknek hívom), az *f*-et a nem elfogadható paraméterkombinációk kiszűrésére használjuk (*f* ilyen értelemben tekinthetjük osztályozó függvénynek is).

Könnyen bővíthető az egy változóra vonatkozó f több szűrő-függvény támogatására. Ha m különböző szűrőfüggvényt szeretnénk alkalmazni (f_i , $i \in 1, ..., m$), akkor az együttes f-et a következőképpen definiálhatjuk:

$$f(x) = \prod_{i=1}^{n} f_i(x)$$
 (11)

²⁵ *A* " ° " a függvénykompozíciót jelöli

Itt a szorzás logikai "és" művelettel analóg. Akkor és csak is akkor lesz *f* értéke 1, valamilyen x paraméterkombináció esetén, ha minden rész-szűrőfüggvény (f_i) az 1 értéket veszi fel a paraméterkombinációra.

Bár az osztályozó segíthet kiszűrni az irreális szimulációkat, a "jó" szimulációk számának aránya (azon szimulációk aránya, amelyek paraméterértékei elfogadhatók voltak) az összes szimulációhoz képest (c_r) egyes esetekben alacsony lehet. Ezt az arányt sikerességi aránynak hívjuk, és a következőképpen határozzuk meg:

$$c_r = \frac{\#sikeres szimulációk}{\#\"osszes szimuláció}$$
(12)

A hagyományos kalibrálási módszerek esetén a poszterior eloszlás által meghatározott paraméterintervallumok sok olyan paraméterértéket tartalmazhatnak, amelyek "rossz okokból jó eredményekre" vezetnek. Ez azt jelenti, hogy nem használhatjuk a teljes eloszlást tudományos ismeretek szerzésére, mivel nem bízhatunk abban, hogy valósághű és stabil eredményeket kapunk. Szükségünk van a paraméterértékeink további vizsgálatára és annak megértésére, hogy miért kapunk sok nem elfogadható eredményt (megjegyezzük, hogy innentől kezdve a nem elfogadható eredményekhez kapcsolódó paraméterértékeket valótlan paraméterértékeknek nevezzük). Ilyen esetekben a felhasználónak meg kell keresnie az a priori intervallumok beállításának esetleges hibáit, vagy több szimulációt kell lefuttatnia, hogy több elfogadható kimeneti értéket kapjon.

Jelen tanulmányban az $f \circ M$ modellt a bemeneti paramétertér egy "fekete doboz" (angolul black-box, abban az értelemben, hogy nem átlátható, azaz bonyolult) osztályozójának tekintjük, ami a bemutatott algoritmus egyik legalapvetőbb felismerése. $f \circ M$ -et ekkor egy white-box (a black-box ellentéte, egyszerű, transzparens) osztályozóval (*g*) közelítjük, olyan módon, hogy a white-box osztályozó által hozott döntések lehető leginkább megegyezzenek a black-box osztályozó által szolgáltatott döntésekkel. Így *g* segítségével módosíthatjuk az a priori paraméterintervallumokat, hogy magas c_r értéket érjünk el.

A legegyszerűbb, mégis rugalmas white-box osztályozók a döntési fák. Döntési fákat széles körben használnak mély neurális hálózatok, vagy support vector machine black-box modellek értelmezése kapcsán (Di Castro és Bertini, 2019; Lee és Kim, 2016). Az osztályozásra használt döntési fák képesek a paramétertér és a kimeneti kategóriák közötti összetett kapcsolatok megragadására (azaz megkülönböztetik az elfogadható és a nem elfogadható paraméterértékeket), miközben ellentétben egyéb egyszerű white-box modellekkel (pl. lineáris modellek), nem túl

érzékenyek a kolinearitásokra²⁶. A döntési fák továbbá segítséget nyújthatnak a paraméterek közötti összetett összefüggések megértésében is.

Egy döntési fa legfelső szintjét gyökércsomópontnak nevezzük. Ez a legfelső szint mindig a legfontosabb paramétert mutatja, amelynek a legnagyobb szerepe van abban, hogy egy modell-futás elfogadható legyen. A fa alsó része különböző rétegekre (vagy szintekre) van osztva, ahol a belső csomópontok találhatók. A különböző belső csomópontokhoz tartozó paraméterek fontossága a rétegek sorszámával csökken (azaz a fontosság fentről lefelé haladva csökken). A belső csomópontok a paraméterértéket feltáró további döntéseket jelentenek, amelyek a paraméter-tartományt két részre osztják az adott szimulációk elfogadható/nem elfogadható jellegétől függően.

A döntési fák elkészítéséhez több módszer állt rendelkezésre (pl.: ID3, C4.5, C5.0, CART; James et al., 2013). Ezek közül – elsősorban egyszerűsége miatt– az úgynevezett CART algoritmust választottam. A döntési fák létrehozása után a pontosságot (*Accuracy*; rövidítve *A*) *a* következő módon határoztam meg:

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$
(13)

ahol *TP* a valódi pozitív esetek számát, *FN* a hamis negatív esetek számát, FP a hamis pozitívként azonosított esetek számát, míg TN a valódi negatív döntések számát jelenti a $f \circ M$ döntési fákkal történő közelítése során.

Amennyiben a pontosság 0,5 alatti (A<0,5), a modell teljesítménye rosszabb volt, mint a véletlenszerű döntésé. 0,7 alatti pontosság²⁷ (A<0,7) esetén az algoritmust nem tekintettem alkalmazhatónak paraméter-intervallum változásra. Ha a pontosság magas, akkor megbízhatunk a white-box modellünkben, mivel hasonlóan viselkedik, mint a black-box.

A döntési fa levélcsomópontjai $(y_1, y_2, ..., y_b)$ folytonos tartományok $(R_1, R_2, ..., R_b)$ a megfelelő döntési csomópontok által meghatározott paramétertérben. Az algoritmus során szeretnénk az eredeti paramétertartományokat olyan módon módosítani, hogy azok a lehető legtöbb elfogadható paramétert tartalmazzák. A legtöbb elfogadható paraméterre vezető "döntési útvonal" tartománya (R_c) az a tartomány, amely alapján az eredeti paraméterteret szűkíteni fogjuk. Mivel a módszer egyik célja az eredeti paraméter-intervallumok olyan módon történő leszűrése, hogy az azok által határolt paramétertérben minél több elfogadható paraméter legyen, R_c -t tovább egyszerűsítjük a legnagyobb

²⁶ Például olyan esetekben, amikor a paraméterek korrelálnak egymással.

²⁷ Hasonlóan más statisztikában használt értékekhez (*p*-érték küszöb) ez is önkényes. A következtetéseink mindig feltételesek. A gépi tanulásban ezt az arányt szokták alkalmazni.

bennfoglalt intervallumok által definiált paramétertérrel (hipertégla). Az így létrejött R_u tartomány a feltételes intervallumcsökkentő eljárás egyik eredménye (10. ábra).



10. ábra. A döntési fa alapú frissítő algoritmus áttekintése egy kétdimenziós paramétertér példáján.
 R_o az eredeti paramétertér, amelyet az [a, b], valamint a [c, d] paraméterintervallumok határozni. A korlátozások alkalmazása után a döntési fa meghatározza az R_c korlátozási tartományt (szürke tartomány), amelyet tovább szűkítettünk a legnagyobb bennfoglalt téglalap alakú régióra (R_u). A frissített intervallumok ez alapján θ₁-re az [a', b'], θ₂-re [c', d']

Ha egynél több olyan terület van, ahol hasonlóan magas az elfogadható paraméterek aránya²⁸, akkor 3 lehetőségünk van:

- 1. Válasszuk ki azt a régiót, amelyikben a legkevesebb összekapcsolt döntési csomópont van, és szűkítsük tovább a paraméterteret eszerint.
- 2. Végezzünk validációt és válasszuk ki azt a területet, amelyik a legkevesebb hamis pozitív eredményt adta és szűkítsük tovább a paraméterteret eszerint.
- 3. Végezzük el a szűkítést az összes különálló területre az eredeti intervallum nem szűkíthető egyértelműen (konkáv összefüggések).

²⁸ A tanulmányban bemutatott példa esetén ez a probléma nem merült fel.

A feltételes tartománycsökkentő eljárás ezt követően igény szerint megismételhető, amennyiben a sikerességi arány nem éri el a kívánt szintet. Az eljárásban a tartomány méretének csökkenésével az adott Monte Carlo szimulációk jobban reprezentálják a teret, így az algoritmus hatásfokának növekedésével lehet számolni (amennyiben a tanítás során figyelembe vesszük, hogy a tanítás viszont egyre kiegyensúlyozottabb lesz).

3.6.1.3. Modell optimalizálás

A döntési fákkal és szűrőfüggvényekkel történő intervallumszűkítés célja olyan, az eredetinél szűkebb paraméter-intervallumok előállítása, amelyben az elfogadható paraméterek aránya maximális. Ez idáig azonban modell-optimalizálás nem történt. Ebben a kontextusban a fent leírt eljárás a modell-kalibráció előfeldolgozó lépéseként is értelmezhető, ahol gépi tanulás segítségével állítjuk elő a kiindulási paraméter-intervallumokat (általánosabban: a priori eloszlását). Az ilyen módon optimalizációval együtt alkalmazott, szűrőfüggvények és döntési fák segítségével végrehajtott tartománycsökkentést feltételes tartománycsökkentő eljárásnak neveztem el (angolul: *Conditional Interval Reduction Method* — CIRM)

A dolgozatomban modell-optimalizálás céljából a hagyományos Bayes-módszerekkel összehasonlítva jóval egyszerűbb GLUE módszert (Prihodko et al., 2008; Stedinger et al., 2008; Beven és Binley, 2014) használtam. A Monte Carlo paramétermintázást a paraméterekre vonatkozó összefüggőségek felhasználásával (ld. Hidy et al., 2021) a Hit and Run algoritmus alkalmazásával végeztem el (lásd 3.2.2. fejezet; Meersche et al., 2009), majd likelihood függvénynek normáliseloszlást használtam²⁹.

Az 1. algoritmus a javasolt feltételes tartománycsökkentő eljárást mutatja be. A munkafolyamat összefoglalja a fent részletezett módszereket, és feltárja az egymást követő lépéseket az összes kapcsolódó bemeneti és kimeneti adattal. Ebben a tanulmányban demonstrációs céllal minden egyes iterációs lépésben GLUE-t végeztem. Az algoritmus kilépési feltételének használt maximális 10 lépést pedig demonstrációs céllal választottam (a maximális iterációs számon túl használhattam volna még a sikerességi arányt is kilépési küszöbként; itt demonstrációs céllal szándékosan magas iterációs számot állítottam be).

²⁹ A modellezési bizonytalanságokat elhanyagolhatónak vettem, a megfigyeléseket pedig függetlennek egymástól.

3.6.1.4. A javasolt módszer összefoglalása

Bemenet	\mathcal{D} , \mathcal{M} , \mathcal{S} , E , G , \overline{e} , \overline{h} , \mathcal{L} , f , N_s , $k = 1, N_i$
1. lépés.	Mintavételezés a konvex paramétertérben: $\theta_i \in S$, majd $f \circ \mathcal{M}(\theta_i)$ kiértékelése
	$(y_i \ keletkezik), \mathcal{L}(d \in \mathcal{D} \theta_i)$ likelihood értékek meghatározása $i \in \{1,, N_s\}$
1a. lépés.	Hit and Run mintavételezés a következőkért N_s megvalósítható $\theta_i \in S, i \in$
	$\{1, \dots, N_s\}$ amelyekre
	$E heta=\overline{e}$
	$G heta\leq\overline{h}$
1b lépés.	Modellszimulációk futtatása, likelihood értékek meghatározása.
1	$o_i = \mathcal{M}(\theta_i)$
	$l_i = \mathcal{L}(o_i, \mathcal{D})$
1c lépés.	Szűrőfüggvény (f) kiértékelése:
_	$y_i = f(o_i)$
Kimenet	Mintavételezett paraméterek és azok y_i kategóriák (0 vagy 1), és a megfelelő like-
	lihood értékek
2. lépés.	Számítsa ki sikerességi arányt (c_r) a (11.) egyenlet segítségével
3. lépés.	Döntési fák képzése az összes kimeneti korlátozó függvényre és az összes y_i -re
4. lépés.	A paraméterintervallumok frissítése a döntési fák alapján, szekvenciálisan
Bemenet	Jelenlegi paraméterek intervallumai, döntési fa
4a. lépés.	Keresse meg azt a levélcsomópontot, amely a legtöbb sikeres esetet tartalmazza a
	képzési halmazból. Ha több jelölt is van, válassza ki a legkisebb számú döntési cso-
	mópont által meghatározott levélcsomópontot, és részesítse előnyben azt a levélcso-
	mópontot, amelynek a legalacsonyabb a hamis pozitív arány (FPR).
4b. lépés.	Módosítsa a paramétertartományokat az előző lépésben kiválasztott csomópontot
	meghatározó döntéseknek megfelelően, ha lehetséges.
Kimenet	Frissített paraméterintervallumok
5. lépés.	Ha $k = N_i$, lépjen a 6. lépésre. Számítsa ki a sikerességi arányt. c_r ; ha c_r elég magas
	ugorjon a 6. lépésre, ellenkező esetben ugorjon az 1. lépésre a módosított interval-
	lummal, és növeljük a <i>k</i> számlálót 1-gyel.
6. lépés.	Végezze el a GLUE vagy más optimalizálási módszert a módosított paramétertarto-
	mányok felhasználásával.
Kimenet	Frissitett parameterintervallumok, GLUE optimum, c_r

l <u>. Algoritmus. A javasolt módszer munkafolyamata. A szimbólumok jelentése a szövegben található.</u>
--

3.6.1.5. Az optimalizációs módszer megvalósítása

Az *a priori*, kukoricára vonaktozó ökofiziológiai paraméterezésének megalkotásához először irodalmi kutatást végeztünk a kukoricával kapcsolatos adatok, például a fajlagos levélfelület, a maximális sztómavezetés, a fény-extinkciós együtthatója stb. után (White et al., 2000). Az empirikus paraméterek (Hidy et al., 2021) esetében az optimalizálást többlépcsős eljárással végeztük. A paraméterek beállítását a németországi Klingenberg termőterületről (DE-Kli FluxNet kód, 50°53'35"N; 13°31'20.6"E) származó eddy kovariancia adatok (bruttó elsődleges produkció (GPP) és evapotranszspiráció (ET)) alapján végeztük (Prescher et al., 2010) a kukorica évek (2007, 2012 és 2018) alapján. Ezen kívül a levélfelület-index (LAI) adatokat is felhasználtuk a helyszínről azokban az években, amikor azok rendelkezésre álltak. A fenológiai fázistól függő allokációs paramétereket kézi (trial and error) módon állítottuk be. A Penman-Monteith-egyenleten alapuló evapotranszspirációs rutin paramétereit szintén a Klingenbergből származó eddy kovariancia ET-adatok felhasználásával állították be.

A korábbi modellezési munkák részeként a kukorica paraméterezését finomították, adatokban gazdag USA kísérleti helyszínek alapján (Bushland liziméteres helyszín Texasban és Mead eddy kovariancia helyszín Nebraskában (US-Ne2 és US-Ne3 FluxNet kódok)). Ez azt jelenti, hogy a kukorica paraméterezését különböző éghajlati viszonyokkal, gazdálkodási típusokkal és különböző kukoricafajtákkal rendelkező helyszíneken tesztelték, amelyeket a FAO-számok sokfélesége jellemez. A modell optimalizálására/validálására irányuló erőfeszítések során kiderült, hogy nem lehetséges egyetlen, univerzálisan alkalmazható kukorica-paraméterezés megalkotása, amely különböző éghajlati és agro-gazdálkodási körülmények között használható. Néhány paraméter erősen helyfüggő (és néhány esetben évjárat függő) volt (a levél elhalásával kapcsolatos paramétere, az allokációs paraméterek és a növényi szövetek élettartamával kapcsolatos paraméterek). Ezen korábbi tapasztalatok alapján készült a modell a priori paraméterezése.

Sajnos jelenleg Magyarországról nem állnak rendelkezésre kukoricával kapcsolatos eddy kovariancia-adatok, annak ellenére, hogy jelenleg 3 helyszínen a szántóföldek felett folyik a mérés. Ez azt jelenti, hogy a modell optimalizálására nagy szükség van más rendelkezésre álló adatok alapján.

A fent javasolt módszert a Biome-BGCMuSo v6.3 optimalizálására használtuk a kukoricára vonatkozóan egy alacsony adatmennyiségű helyzetben (ami azt jelenti, hogy kevés jó minőségű megfigyelési adat áll rendelkezésre), amikor csak a terméshozamra vonatkozó végleges adatok álltak rendelkezésre, kiegészítve néhány további információval a termés általános tulajdonságairól (amelyek korlátokat jelentenek). A Biome-BGCMuSo-t parcellaszinten futtattuk Martonvásáron egyetlen általános kukoricaparcellát szimulálva, 220 kg N/hektár április elején kijuttatott műtrágya mennyiséggel. A vetési és betakarítási időpontokat a kísétler beállításában használt időpontoknak megfelelően határoztuk meg. A meghajtó meteorológiai adatokat a FORESEE adatbázisból nyertük (lásd 3.4. alfejezet).

A modell bemeneti talajfájlját a talaj textúrájára vonatkozó megfigyelések és a talaj vízvisszatartási görbéjének mérései alapján állították össze Martonvásárra, amelyekhez a pipettás módszert (ISO 11277) alkalmazták a bolygatott, illetve a homok/kaolinos dobozos módszert (ISO 11274) a 100 cm³-es bolygatatlan magmintákon, amelyeket az LTFE-kből gyűjtöttek. A nitrogénciklus legtöbb paraméterére a Hidy et al. (2021) által javasolt értékeket használtuk.

A megfigyelési pontra vonatkozó optimalizáció eredményét megyei szintű termásátlagokkal validáltuk. A megyei szintű szimulációkhoz is a FORESEE adatbázis szolgáltatta a meteorológiai

adatokat. A DOSoReMI adatbázis (Pásztor et al., 2020) volt a rácshálós talajadatok forrása (lásd 3.5. fejezet). A Fejér megyére vonatkozó szimulációkat egy előre meghatározott 0,1°× 0,1° felbontású rácson végeztük el. A műtrágya mennyiségét a Központi Statisztikai Hivatal adatai alapján állítottuk be. A vetési idő április 15., a betakarítás időpontja pedig október 10. volt minden szimulációban. 51 rácscellát használtunk, amelyek 4358 km-es területet fedtek le. A szimulált termésadatokat megyei szinten aggregáltuk a cellaspecifikus terméseredmények egyszerű, évenkénti átlagolásával.

A Biome-BGCMuSo 120 kukoricával kapcsolatos ökofiziológiai paramétert tartalmaz, amelyeket a modellezőnek kell beállítania a szimulációk előtt. A paraméterek egy része általános növényi paraméter (mint például a növényi részek C:N aránya; maximális sztómavezetés; maximális gyökérzetmélység; gyökéreloszlási paraméter stb.), míg más részük kukoricaspecifikus (pl. az virágzás alatti hőstresszt, a csírázást a talaj víztartalmának függvényében befolyásoló paraméterek stb.; Hidy et al., 2021). A 120 paraméter közül 42-re van hatással néhány szabály. A 42 szabály által befolyásolt paraméterből 28 az allokációval kapcsolatos.

Hasonlóan más, nagyszámú paraméterrel jellemezhető modellekhez (pl. Bilionis et al., 2015), lehetetlen a modellt minden paraméterre optimalizálni. Ehelyett a paraméterek kiválasztásához és a paraméterek beállításához korábbi tapasztalatokat és több korábbi modellértékelés eredményét használtuk fel. A korábbi modell-optimalizálási erőfeszítések során (ld. Módszerek) a legfontosabb paraméterértékek kiválasztása részben objektív érzékenységi elemzéssel, részben pedig manuális paraméterbeállítással történt. Számos paramétert a rendelkezésre álló megfigyelési adatok (mint például a levél, a szár és a gyökérzet C:N aránya), valamint a német és amerikai kísérleti helyszíneken szerzett tapasztalatok alapján határoztunk meg (maximális sztómavezetés, lomkorona fényelnyelési együtthatója stb.). A fennmaradó paraméterek a helyszínek között változónak bizonyultak, és a modell érzékenységet mutatott a paraméterek megfelelő beállítására (2. táblázat). Ezeket a paramétereket választottuk ki optimalizálásra.

A kiválasztott paraméterek esetében a modelloptimalizálás előtt a felső és alsó határokat a szakirodalmi értékek alapján határoztuk meg (White et al., 2000). Fodor Nándor és Árendás Tamás szakértői tudását és korábbi tapasztalatait is felhasználtuk az intervallumok beállításához (2. táblázat). A kukorica szimulációinak teljes a priori ökofiziológiai paraméterezését az M1 mellékletben mutatom be.

2. táblázat. Az optimalizálásra kiválasztott modellparaméterek teljes listája az előzetes paramétertartományokkal. A bevezetett jelölésrendszer szerint (3.6.1.1. alfejezet) MIN az I₁, a MAX az I₂.

	<i>y</i>		,2'
Rövidítés	Leírás	MIN	MAX
Rubisco	A levél teljes nitrogén-tartalmának százalékos aránya a	0,07	0,12
	Rubisco enzimben. Szabályozza a fotoszintézis maxi-		
	mális értékét (White et al., 2000).		
root_distribution_paraméter	Empirikus paraméter a gyökerek talajrétegeken belüli eloszlásának kiszámítására (Jarvis, 1989).	2	6
root_we-	Ez a paramétert a növények empirikus gyökerezési	0,08	0,14
ight_to_max_root_depth	mélységét szabályozza.		
root_depth_function_shape	Ezt a paramétert a növények empirikus gyökérmélység- számításánál használják (a 4M-modell módszere alap- ján). A paraméter a gyökérmélység időbeli profilját sza- bályozza (lehet konvex vagy konkáv az idő függvényé- ben).	0,4	1,6
senescence_coeff_for_leaf	A talajnedvesség-stresszhez kapcsolódó mortalitási együttható, amely a levélelhalás mértékét szabályozza (a hosszan tartó szárazság okozta stressz miatt egy nap alatt elhaló levélszövetek aránya).	0,001	0,04
vízstressz hatása a fotoszin- tézisre	Empirikus paraméter, amely szabályozza a nem sztóma eredetű talajvíztartalom stressz fotoszintézisre gyako- rolt hatását (azaz az asszimiláció leszabályozását).	0	0,7
length_of_phenophase_3	A 3. fenofázis hossza hőfok-napokban (GDD) kife- jezve. A kukorica esetében ez a korai vegetatív növeke- désre vonatkozik.	240	450
length_of_phenophase_4	A 4. fenofázis hossza GDD-ben kifejezve. A kukorica esetében ez a késői vegetatív növekedésre vonatkozik (ez a fenofázis a virágzással ér véget).	240	450
length_of_phenophase_6	A 6. fenofázis hossza GDD-kben kifejezve. A kukorica esetében ez a virágzást követő szemtelítődésre vonat-kozik.	850	1200
leaf_allocation_3	A teljes napi allokáció levélnövekedéshez kapcsolódó része a 3. fenofázisban.	0,4	0,5
root_allocation_3	A teljes napi allokáció gyökérnövekedéshez kapcso- lódó része a 3. fenofázisban.	0,3	0,5
stem_allocation_3	A teljes napi allokáció szárnövekedéshez kapcsolódó része a 3. fenofázisban.	0,2	0,5
leaf_allocation_4	A teljes napi allokóció levélnövekedéshez kapcsolódó része a 4. fenofázisban.	0,2	0,5
root_allocation_4	A teljes napi kiosztás gyökérnövekedéshez kapcsolódó hányada a 4. fenofázisban.	0,2	0,4
stem_allocation_4	A teljes napi allokáció szárnövekedéshez kapcsolódó része a 4. fenofázisban.	0,3	0,5
root_allocation_6	A teljes napi allokáció gyökérnövekedéshez kapcso- lódó része a 6. fenofázisban.	0,05	0,15
fruit_allocation_6	A teljes napi allokáció azon része, amely a 6. fenofázis- ban a szemtelítődéshez kapcsolódik.	0,5	0,8
stem_allocation_6	A teljes napi allokáció szárnövekedéshez kapcsolódó hányada a 6. fenofázisban.	0,1	0,4
maxlifetime_3	Az új levélszövet maximális, genetikailag meghatáro- zott élettartama a 3. fenofázisban, GDD-ben kifejezve.	500	1300
maxlifetime_4	Az új levélszövet maximális, genetikailag meghatáro- zott élettartama a 4. fenofázisban, GDD-ben kifejezve.	500	1300

Az optimalizáláshoz normál likelihood függvényt használtunk (1. egyenlet), és feltételeztük, hogy a modellezési bizonytalanságok elhanyagolhatóak a megfigyelési bizonytalanságokhoz képest. A white-box modell a fent leírt CART-tal ellátott döntési fa volt. A döntési fák komplexitáson alapuló metszési tényezőjeként 0,01-et használtunk, mivel ez az *rpart* csomag alapértelmezett értéke (Therneau et al., 2022)

Négy szűrőfüggvényt használtunk (*f*). Az első az éves betakarítási indexhez (HI) kapcsolódik, amelyet a végső szemtermés és a betakarításkori teljes föld feletti biomassza aránya határoz meg (Goudriaan et al., 2001). A szimulált HI-értékek mediánjának a több tudományos publikáció alapján meghatározott [0,40, 0,55] tartományban kell lennie (Ion et al., 2015; Li et al., 2015; Hütsch és Schubert 2017; Liu et al., 2020). Az éves maximális levélfelület-index (LAI_{max}) mediánját 2,7 és 5 m²/m² közé becsültük, ami egy megfigyelésen alapuló beállítás Magyarországra (lásd pl. Pokovai és Fodor, 2019). A gyökérzetmélység hosszú távú mediánértékének a virágzási fenofázis kezdetén 1,4 m-nél nagyobbnak, de 1,8 m-nél kisebbnek kell lennie (ezek az értékek szakértői ismereteken alapulnak). A virágzási idő mediánjának (az év napjában kifejezve: DOY) 180 és 190 között kell lennie, ami a Magyarországra jellemző tartomány.

A szűrő függvényeket a modell eredményei alapján egy egyszerű algoritmus határozza meg. Például, ha az éves LAI_{max} egy vektor, amely a Biome-BGCMuSo szimulált LAI éves maximális értékeit tartalmazza, akkor *f* a következőképpen definiálható:

$$f(x) = \begin{cases} 1, & \text{med}(LAI_{max}) \in [2.7,5] \\ 0, & \text{máskülönben} \end{cases}$$
(14)

ahol med a szimulált évek értékeinek mediánját jelenti a pont-szimuláció során.

A jelen dolgozatomban bemutatott esettanulmányhoz az RBBGCMuso CIRM ágát (branchét) használtam. A kimeneti kondicionálási módszert 10 iteratív lépéssel valósítottam meg. Minden egyes iterációs lépésnél 10 000 szimulációt végeztünk. A munka során többnyire a standard R programot használtuk (R Core Team, 2021). Emellett az rpart csomagot használtuk a döntési fák konstruálásához (Therneau et al., 2022).

Minden egyes iterációs lépésben a maximális valószínűségű paraméterkészletet tároltuk. Az poszterior paraméterintervallumokat az előre meghatározott kondicionálásnak "jól viselkedő" (felső 5%) paraméterértékek alapján becsültük. Az első 9 iterációs lépés során a paraméterintervallumok csökkentése a döntési fa frissítési algoritmusa alapján történt. A GLUE-alapú poszterior intervallumcsökkentést csak az utolsó iterációs lépésben vettük figyelembe. Az optimális paraméterkészletet a "jól viselkedő" paraméterek mediánjaként számoltuk ki. Az alábbiakban ezeket az intervallumokat és az optimális paraméterkészletet GLUE-alapú, korlátozott intervallumoknak nevezzük.

Az optimalizált modell teljesítményét a lineáris korrelációs együttható négyzete (R²), a torzítás (szisztematikus hiba), az átlagos négyzetes hiba (RMSE) és a Nash-Sutcliffe-féle modellezési

hatékonyság (ME) segítségével értékeltem (Ma et al., 2011; Sándor et al., 2016). A modell jóságát jellemző mérőszámokat a megfigyelt és a szimulált kukoricatermés idősorok alapján számítottam ki. Az optimalizált modellt a gyakorló adathalmaz alapján, valamint a megyei szintű szimuláción is értékeltük független adatok felhasználásával. Ez utóbbi esetben az átlagos éves kukoricatermést a modellszimulációkból számítottam ki, és a végleges idősorokat a megfigyelt KSH adatokkal összevetve értékeltem. A végleges (GLUE) paraméter-intervallumok változásának százalékos arányát is számszerűsítettem az a priori intervallumokhoz képest.

3.7. A modell térbeli kiterjesztését támogató keretrendszer: AgroMo

Az alábbiakban az AgroMo keretrendszert mutatom be, részletezve a tervezésekor használt logikát és a főbb informatikai megoldásokat. Az AgroMo egy felhasználóbarát grafikus felületet, ami a modell-adat fúzió majdnem teljeskörű megvalósítását tartalmazza, valamint a magas szintű modellalkalmazáshoz szükséges keretrendszert, és a hatásvizsgálatokhoz szükséges funkcionalitást nyújtja.

3.7.1. AgroMo programozási háttere

Az RBBGCMuso-ra épített AgroMo magja szintén R-ben íródott, segítve az RBBGCMuso funkcióinak szabadabb elérését rendszerfolyamatok közötti kommunikáció nélkül. Ilyenformán az RBBGCMuso az AgroMo csomagfüggősége.

Hasonlóan az RBBGCMuso-hoz, az AgroMo fejlesztéséhez is a korszerű, Linus Torvalds által kifejlesztett verziókövető rendszert, a Git-et használtam, amely segítségével a fejlesztési folyamat átláthatóvá és biztonságossá tehető, bármely régebbi verzióra vissza lehet lépni, alternatív fejlesztési utakat/ágakat (brancheket) lehet létrehozni. Ezekben a fejlesztési ágban különböző funkciók is biztonságosabban fejleszthetők anélkül, hogy a már jól működő eredeti fejlesztési ágba véletlen hibákat ejtenénk a programozás során. Én a következő fejlesztési ágakat hoztam létre: *master* (stabil kód, a felhasználók ezt használják), *devel* (fejlesztési kód, a különböző funkciók tesztelése után ide kerülnek be a kódok, itt újabb tesztelésen esnek át; amennyiben kellő stabilitású a devel ág, a változásokat a masterbe fűzzük), *enhancement* (különböző fejlesztéseket tartalmazó ideiglenes ágak), *bugfix* (hibajavító ágak), *docs* (dokumentáció), GUI (a név a Graphical User Interface angol rövidítése után lett elnevezve, amelyben gyakorlati okokból külön ágat kaptak a grafikus felhasználói felületeket érintő fejlesztések, amelyek többnyire a HTML elemek stílusának Cascading Style Sheet-en (CSS-en; weboldalak stílusát szabályozó nyelv) keresztül való szabályozását jelentette).

Git szervernek a GitHub-ot használtam, mivel ez azon túl, hogy megbízható tárhelyet biztosít, egyéb projekt-management funkciókat is lehetővé tett. Feladatokat lehetet kijelölni, hibákat,

kérdéseket lehet felvenni a felhasználótól, mérföldköveket tudunk definiálni, dokumentációs lehetőségei is vannak (README.md megjelenítése weboldalként). A GitHub többféle licenszt támogat, így lehetőségünk volt a szoftvert nyílt forráskódú, "szabad" formában, GPL-3 COPYLEFT licenszben közzétenni, ami bárkinek megengedi a szoftver letöltését, módosítását, terjesztését azzal a feltétellel, hogy a licensz, a szerzői és a használati jogok megmaradnak.

Ezen licensz lehetővé teszi, hogy a csomag megfeleljen az R általánosan elérhető csomagtárjában, a CRAN-ban, így kapva jóval nagyobb publicitást. Egy ilyen csomag-publikációnak viszont sok feltétele van a csomag minőségére vonatkozóan, s a feltöltést követően lektorálják is. A lektorálási folyamatot számos csomag hivatott egyszerűbbé tenni, amelyekből kettőt használtam a fejlesztéshez: a devtools (Wickham et al., 2022) és a roxygen2 (Wickham et al., 2022). Az első a saját készítésű R csomagok fejlesztését, a másik a fejlesztés dokumentálását segíti. Ezzel a forráskódban egyidejűleg van jelen a programkód és a dokumentáció. Az amúgy nehezen használható LaTeX-re emlékeztető beépített dokumentációs nyelvét helyettesíti az eszköz.

Mivel az AgroMo-t úgy terveztük, hogy az grafikus felületet szolgáltasson és tartalmazza azokat a futási logikákat, amelyek egy rács-alapú mezőgazdasági keretrendszertől elvárhatóak, szükség volt egy grafikus könyvtárra (csomagra), amely az ilyen irányú fejlesztéseket támogatja. Itt több lehetőségünk volt: tcl/tk, RGTK, RQt, Shiny (Chang et al., 2017). Ezek közül a Shiny-t választottam, mert ezzel web-alapú alkalmazást tudunk készíteni, amelyet minden olyan eszközön el tudunk érni, amelyen modern böngésző elérhető, továbbá így lehetőségünk nyílik az interaktív grafikonok készítésére a plotly csomagon (Sievert et al., 2017) keresztül.

Az AgroMo forráskódja a <u>https://github.com/hollorol/AgroMo</u> honlapon elérhető. A fejlesztési módszertant és lépéseket a projekt szerkezete továbbá az úgy nevezett "commit message"ek. Minden logikailag elválasztható munkalépéshez tartozó egyszerű üzenetet giten keresztül a "git log" parancs segítségével meg lehet tekinteni.

3.7.2. A modell térbeli kiterjesztésének lehetőségei

Mivel a Biome-BGCMuSo folyamatai térben, horizontálisan nem explicitek, a futtatások mindig homogén pixelt feltételeznek. Erre a limitációra talán a legegyszerűbb megoldás, hogy minden homogénnek feltételezett térbeli egységben (innentől cellában) futtatjuk a modellt, és a cellánkénti eredményekből tetszőleges statisztikával, vagy akár a teljes adatsorral jellemezhetjük a területet. Például lehetőségünk nyílik térbeni mintázatok, szabályszerűségek felfedezése, de ha szeretnénk, akkor területi átlagokat is számolhatunk. Ez a megoldás jóval valósághűbb eredményeket ad, mintha egyszerűen egyetlen a futtatással próbálnánk a teljes területet jellemezni.

Az alkalmazott 0,1×0,1 fokos térbeli felbontású rácsháló összhangban van a FORESEE adatbázisának geometriájával, és az ország határai mentén odafigyeltünk arra, hogy a létrehozott rendszer a lehető legkisebb területű tartalmazó háló legyen, amelyben a magyarországi rész a négyzet teljes területének legalább a felét teszi ki (11. ábra). Ez együtt járt azzal, hogy 93 000 km² terület helyett 110 400 km²-nyi területet fedtünk le.



11. ábra. Az általunk definiált AgroMo rácsháló, amely a teljes országot lefedi.

3.7.3. AgroMo adatbázis és fájlrendszere

A modell térbeli kiterjesztése során alkalmazott rácshálós megoldás adatigénye jelentős. Minél finomabb a rácsfelbontás annál nagyobb az adatigény, hisz a növekvő cellák számával arányosan több input fájlra van szükségünk: meteorológiai-, talajadatokra, valamint művelést meghatározó fájlokra is. A modell futtatásának következményeként pedig számos olyan fájl is keletkezik, amelyek a modell kimeneteit tárolják. Ha olyan rendszert szeretnénk alkotni, amelyet informatikai tudás nélkül is hatékonyan tudnak a felhasználók használni, akkor ezeket az adatokat mindenképp szükséges valamilyen módon strukturálni. Erre kínál megoldást, mint módszertani keret a relációs algebra (Codd, 1970). Az adatot ebben a változók (attribútumok) és relációik (attribútumok közti kapcsolatok) írják le, amelyeket a táblázatok rendszereként érdemes elgondolni, ahol a változók közti kapcsolatokat megkötésekkel lehet biztosítani. A relációs adatmodell a mai napig elterjedt, mivel segítségével garantálható a relációk és a tranzakciók (változások az eltárolt adatokban) integritása. Ehhez az adatmodellhez alkották meg az SQL (Structured Query Language) nyelvet is, amellyel az emberi nyelvhez hasonló mondatokban kérhetünk le adatokat, valamint végezhetünk analitikai műveleteket. A relációs adatbázisokban történő tárolásnál több szempontot kell egyszerre figyelembe venni. Ilyen lehet az adatbázis tárhely igénye, a redundanciák elkerülése, a beillesztések, adatfeltöltések sebessége, illetve az adatintegritás biztosítása, továbbá a lekérdezések sebessége. Mivel a bemeneti adatok és a keletkezett adatok is többféle módon vannak tárolva, az elsődleges szempont a lekérdezések sebessége volt, mindent ennek rendeltünk alá. Az adatmodellünknek azon részeit, amelyeket gyakran használunk denormalizált módon, redundánsan tároltuk. Erre jó példa a dátumok tárolása, amely esetén ugyanazon táblában egész számként tároltuk az éveket, hónapokat, napokat, valamint a dátumot magát is, aminek tárolására elviekben nem lenne szükségünk, mégis dátum-műveleteket sokkal gyorsabban tudunk dátumon végrehajtani, évek, hónapok szerinti csoportosítást pedig ezen változók segítségével hatékonyabban tudunk végezni, mint a dátumon magán. Egy további gyorsítási lehetőség, hogy index-táblákat tárolunk az adatok mellett. Ezek a táblák olyan célt szolgálnak, mint a fogalom- vagy tartalomjegyzékek a könyvek esetében: sokkal gyorsabban lehet az adatokat megtalálni keresésnél. Nyilvánvaló, hogy ez az adatbázis méretének növekedésével jár, továbbá az is, hogy a betöltést is lelassítja.

Az AgroMo-t egy felhasználós rendszernek terveztük, így a beágyazott rendszerekhez kifejlesztett egy felhasználós SQLite (Hipp, 2020) adatbázist használtuk. Az adatbázis a legalapvetőbb SQL függvényeket támogatja, nincs benne felhasználó-kezelés, az adatbázist egy egyszerű bináris fájlban tárolja. Ez az adatbázisokat hordozhatóvá is teszi. Az AgroMo-n belüli adatbázisokat a 3. táblázat foglalja össze.

0		
	Adatbázisnév	Hely a fájlrendszerben
Klímaadatbázis	weather.db	database/weather.db
Talajadatbázis	soil.db	database/soil.db
Pontfutások eredményeinek adatbázisa	output.db	output/site.db
Rácshálós futások eredményei- nek adatházisa	grid.db	output/grid.db

3. táblázat. AgroMo adatbázisok és helyeik a könyvtárrendszerben

Az adatbázisok létrehozásához és a hatékony (gyors, párhuzamosított) betöltéshez GAWK-t (Free Software Foundation, 2023), valamint GNU Parallelt használtam (Tange, 2015).

A centrális modell viszont kizárólag fájl alapú, így adatbáziskapcsolattal közvetlenül nem vezérelhető. A fájlokat előre kell legenerálni, vagy adatbázisból szkriptek vagy külső szoftver segítségével. Az alkalmazásnak pedig ismernie kell azokat a fájlokat, amelyeket a modell számára szolgáltat. Erre két lehetőség áll fenn: vagy egy globális leírófájlt készítünk és ezt követően a modell számára az ideiglenes mapparendszert dinamikusan, futási időben generáljuk, vagy pedig egy fix könyvtárstruktúrában várjuk az adatokat. Ez utóbbi egyszerűbbnek bizonyult, így saját

könyvtárszerkezet (12. ábra) mellett döntöttünk. Az adatbázisok ebben a könyvtárszerkezetben találhatóak (3. táblázat).



12. ábra. AgroMo fájlrendszer áttekintése.

3.7.4. AgroMo lekérdezésekhez használt metanyelv

A talaj-víz-levegő rendszer ilyen módon történő egységes adatbázisrendszerben való kezelésével lehetőségünk nyílik a különböző komponensek kapcsolatát leíró folyamatok feltárására egyszerű lekérdezésekkel. A relációs adatbázisok a különböző változók (attribútumok) közti relációk feltárására többnyire SQL lekérdező nyelvet használnak. Ehhez viszont szintén szükséges alapszintű adatbázis ismeret, amelyet az AgroMo felhasználóitól nem várunk el. Ehelyett felfedezve, hogy a lekérdező nyelvben alkotott mondatok a természetes nyelvhez hasonló redundanciával bírnak, ahol gyakran egy-egy mondat kizárólag néhány mondatrészben tér el, egy SQL nyelvre épülő saját doménspecifikus metanyelvet alkottam. Például, amennyiben 10 adattáblánk van (termésadatokkal), amelyben az adatok 2009 és 2100 közötti időszakra vonatkoznak, és szeretnénk meghatározni az éves termésátlagok minimumát, maximumát, szórását, terjedelmét, a kiválasztott időszaktól és aggregálási

összefüggéstől függően 180 000 különböző lekérdezést írhatunk³⁰. A metanyelv 9+2 helyettesítést jelölő szimbólumot támogat (általános hely-jelölő: {1}, ..., {9}, időszakleíró: [T1], [T2]). A nyelvnek alapvetően 5 eleme van (13. ábra):

- 1. SQL mondat ("query")
- 2. Természetes nyelven megírt mondat ("Names")
- 3. Helyettesíthető részeket jelölő elemek (Pl. {1}, [T1], T[2]). Ezek a kicserélhető részek az SQL mondat helyettesíthető részeiben is megjelennek.
- a helyettesíthető részek lehetséges megfejelői a természetes nyelvben ("optionAlias").
- 5. a helyettesítő részek megfelelői az SQL mondatban ("options")



13. ábra. Lekérdező metanyelv példa.

 $^{^{30}}$ A [a,b] intervallumból $\binom{b-a+1}{2}$ számú részintervallum választható ki.

Lehetséges a lehetőségek dinamikus értékmegadására az *options*, illetve az *optionAlias* részekben. Ekkor a manuális megadás helyett a ***dinamikus_valtozo*** bejegyzés a *get_dinamikus_valtozo* függvény segítségével automatikusan generálódik. Ennek szükségességét indokolják meteorológiai (*meteorology*/get_meteorology), a kimeneti (*tables*/get_tables), valamint a talajadatbázisok táblái (*soil*/get_soil), amelyek folyamatosan bővülnek.

A háttérben a megfelelő nyelvi átalakításokat – az informatikai világban gyakran alkalmazott – reguláris kifejezések segítségével végeztem (az átalakításra példa a 14. ábra). A grafikus felületen történő változásokat és automatizmusokat pedig saját JavaScript függvényekkel oldottam meg (ld. AgroMo forráskód).

TÉRKÉP | Klímaváltozás (ensemble)HATÁSA A TERMÉSRE CELLÁNKÉNT a(z) [2020-2100] időszakra a(z) {1:agromo} szimuláció alapján

```
SELECT baseline.cell id,
       100* (projection.value - baseline.value) /baseline.value
FROM (
      SELECT cell_id, AVG(value) AS value
      FROM (
             SELECT cell id, year, AVG(TgrainDM) AS value
             FROM agromo
             WHERE year>=2001 AND
                   year<=2020
             GROUP BY year, cell id)
      GROUP BY cell id) AS baseline
INNER JOIN (
      SELECT cell id, AVG(value) AS value
      FROM (SELECT cell id, year, AVG(TgrainDM) AS value
             FROM agromo
             WHERE climate id>10 AND
             year>=2020 AND year<=2100
             GROUP BY year, cell id)
      GROUP BY cell id) AS projection
ON baseline.cell id=projection.cell id;
```

14. ábra. Példa a metanyelv-SQL átalakításra.

3.7.5. AgroMo szcenáriók és domének

A klímaadaptív mezőgazdaság feltétele, hogy a körülményekhez alkalmazkodó növénytermesztési döntéseket hozzunk. Ehhez azonban elengedhetetlen a természetes folyamatok mellett az emberi döntések (vetési idők, öntözés, földhasználat váltás) bizonyos fokú modellezése is. Magyarországon (és Közép-Európában) az éghajlatváltozásra vonatkozó szcenáriók a FORESEE adatbázison keresztül vizsgálhatók (RCP szcenáriók). A növénytermesztést közvetlenül érintő döntéseket viszont ettől függetlenül külön kell modellezni. Fontos, hogy a döntéseknek az időbelin túl, térbeli dimenziója is van. Az ország más és más részeire vonatkozóan eltérő döntéseket hozunk. Azon site-ok halmazát, amelyekre vonatkozóan megegyező döntéseket hozunk doménnek nevezzük. A szcenáriókat (időszakok, domének, döntések) hármasával definiáljuk³¹. Ezen hármas alapján egy külső program (Fodor Nándor eszköze³²) az AgroMo fájlrendszerében létrehozza az összes modell számára fontos input fájlt, továbbá egy forgatókönyv fájlt (storyline fájlnak hívjuk), ami egy egyszerű fájl, aminek a szerkezete a 15. ábrán látható. Ez alapján az AgroMo képes a forgatókönyvek szerint a modelleket lefuttatni, mivel az INI fájl egyértelműen meghatározza a management fájlokat.

Szcenárió név, leírás Napi;változók;pontosveszővel; szeparálva Éves;változók;pontosveszővel; szeparálva INI név;verzió név;kezdő év; befejező év

15. ábra. A storyline file szerkezete.

3.7.6. AgroMo kalibráció

Mint ahogy arról a bevezetésben részletesen is írtam, a modell használhatóságának, azaz jóságának feltétele, hogy a modell megfelelő módon be legyen állítva ("fel legyen paraméterezve"). Az AgroMo-nak ezért a futtatáson túl képesnek kell lennie a paraméterekkel kapcsolatos vizsgálatok egyszerűsített elvégzésére. Ehhez az AgroMo fájlrendszerben az *input* mappával egy szinten a *calibration* mappát használjuk. A *calibration* mappában helyezkednek el a különböző kalibrációs projektek különálló könyvtárai (például 16. ábrán a megyei szintű kalibrációs projektek; lásd Eredmények).

agromoklima[~/desszertacio]∎ tree -L 1 calibration			
calibration			
Martonvasar_maize			
MULTISITE_OBS.xlsx			
NUTS3_wheat			
6 directories, 2 files			

16. ábra. Kalibrációs mappa szerkezete.

³¹ Ezek a jövőre vonatkozó feltételezéseinket foglalják magukba

³² A storyline creator elérhetősége: <u>https://owncloud.atk.hu/index.php/s/IaxVKXIRKgTz54W</u>

3.7.6.1. Set és Obs fájlok

Minden kalibrációs projektben található egy *set* kiterjesztésű fájl, ami tartalmazza a kalibráláshoz felhasznált kimeneti változók nevét pontosvesszővel elválasztva, valamint egy *parameters.csv*-vel (ld. 3.2.1 fejezet) azonos szöveges részt, ami az RBBGCMuso-hoz hasonlóan az apriori paraméter intervallumokat definiálja. A mérési adatokat a kalibráció számára az *obs* kiterjesztésű pontosvessző szeparált fájlok biztosítják. Ezekben fellelhető a kimeneti változó neve (ez kulcsolja össze a mérési adatot a *"set"* fájlban található kimeneti változón keresztül a modell kimenetével), a domain_id (több-cellára vonatkozó kalibráció esetén használjuk csak, egy domén több cella együttese), a mérés dátuma, az átlaga, bizonytalansága, minimum, illetve maximum értéke.

3.7.6.2. Constraints fájl

Ezen kívül a mappa tartalmaz még a *constraints.json* nevű fájt, ami a CIRM (3.6.1 alfejezet) optimalizációnál alkalmazott szűrőfüggvényeket, valamint azok paramétereit tartalmazza. A függvények általános definiálásához saját domain specifikus nyelvet használok, amelyben a "]" a kompozíció operátor³³, a "SELECT" függvény egy magasabb-rendű függvény (angolul "higher-order function"), amely egy adott kimeneti változó éves adataira alkalmaz egy tetszőleges R függvényt (akár lehet névtelen függvény is). A "." a kompozíció során az adatok behelyettesítésének helyét mutatja. Például, a következő kifejezés a kimeneti adatok éves LAI maximumainak mediánját számítja ki:

SELECT(lai, max) | quantile(.,0.5)

A *constrain.json* két objektumból áll: "constraints", valamint "treshold". A "constraints" objektum a hatérértékekből, valamint a kifejezésekből álló objektumok tömbje, a "treshold" pedig egy arányszám, ami arra vonatkozik, hogy a térbeli kalibráció esetén a feltételek az összes cella hányad részére teljesüljenek. A *constraints.json* fájlra egy példa a 17. ábrán látható.

3.7.6.3. Cal fájl

Az AgroMo kalibráció szempontjából legfontosabb fájlja az úgy nevezett *cal*-fájl. Ez tartalmazza a hivatkozást az összes eddigi fájlra (első sor: *obs* fájl, második sor: *set* fájl, harmadik sor: *constraints.json*), valamint a kalibrálás során felhasznált INI fájlok helyére (4. sor), továbbá ebben definiálódnak a kalibráció során alkalmazott domének is (soronként az INI fájlok az INI kiterjesztés nélkül, valamint a domain_id [az observation fájlban használt domain_id-vel megegyező]).

³³ Hasonló eszköz a UNIX(szerű) rendszerek "pipe" operátorához. $(f \circ g)(x) = (g | f)(x) = f(g(x))$



17. ábra. Példa a constraints.json fájlra (a SELECT függvényen belüli változók értelmezéséhez ld. az AgroMo fájlrendszer központi adattábláját (3.7.3 fejezet)).

3.7.7. Disztribúció

Az AgroMo szoftver nemzetközi adaptációjának céljából, a használhatóság és elérhetőség maximalizálására törekedtünk. Kulcsfontosságúnak ítéltük, hogy a telepítési folyamat ne jelentsen technikai kihívást a végfelhasználók számára, ezért az összes szükséges R csomagot egy átlátható mapparendszerbe helyeztünk. A rugalmasság érdekében egy hordozható R környezetet integráltunk a szoftverbe. Az R által elindított shiny alkalmazást az nwjs (https://nwjs.io/) minimalista böngésző jeleníti meg. A felhasználói interakció egyszerűsítése céljából inicializáló szkripteket és batch fájlokat alkalmazunk. Fodor Nándor, a Delphi programozási nyelv használatával, rendszerindító alkalmazást fejlesztett, mely automatikusan aktiválja a szükséges batch fájlokat. Mindezen elemek kombinációja révén az AgroMo egy hordozható, telepítést nem igénylő, natív élményt nyújtó webalkalmazássá vált a Windows operációs rendszeren.

3.8. Klímaváltozás hatástanulmány

Az általam fejlesztett keretrendszer hatékonyságának demonstrálása céljából az éghajlatváltozás hazai kukoricatermelésre gyakorolt várható hatása kapcsán végeztem egy vizsgálatot. Az esettanulmányban a FORESEE adatbázisból 14 éghajlati modellből a CNRM_ALADIN53, a HADGEM2_CCLM, a HADGEM2_RACMO22E, az MPI_CCLM valamint NCC_HIRHAM5 éghajlati modell RCP4.5 és RCP8.5 forgatókönyveit használtam fel. Arra a kérdésre kerestem a választ, hogy közel változatlan termesztési gyakorlat mellett milyen hatással lesz az időjárás és a légköri CO₂ szint változása a termésmennyiségre.

Megfigyelésként a KSH megyei szintű adatait használtam fel. A kezelés kapcsán április 15-i vetésidővel dolgoztam, és október 10-i aratással. A KSH alapján megyénként változó műtrágyázást feltételeztem, átlagosan 150 kgN/ha/év dózissal, április 1-i kijuttatással. Öntözést nem feltételeztem. Mivel közvetlenül a rendszerváltás (1989) után drasztikusan lecsökkent az alkalmazott műtrágya mennyisége országos szinten (Kern et al., 2018), ezért a referencia időszaknak a 2001-2020 időszakot tekintem, amikor már ismét viszonylag magas volt a kijuttatott mennyiség. A referencia időszakhoz képest a közeli jövő változását a 2041-2060 időszak alapján, míg a távoli jövő termésmennyiségének alakulását a 2081-2100 időszak alapján értékelem.

4. Eredmények

4.1. Az RBBGCMuso keretrendszer

Munkám eredményeként az RBBGCMuso szoftvercsomag funkcionalitása jelentősen kibővült, és a GitHub-on egy részletes oktatóanyag is született, ami alapján még a kezdő modellező is viszonylag hamar tud futtatni sikeres szimulációkat (Hollós et al., 2023)

Az RBBGCMuso szoftvercsomag összesen 135 (köztük 64 magas funkcionalitású), súgóval rendelkező függvényből áll. A szoftver laikusok számára is könnyen telepíthető. Mindösszesen egy ingyenesen elérhető R szoftvercsomagra van szükség a telepítéséhez, és Internet kapcsolatra természetesen. R-ből (például a közkedvelt RStudio felületén keresztül) az alábbi parancsokat kell kiadni, és a csomag rövid időn belül települ.

```
> install.packages("remotes")
```

```
> remotes::install github("hollorol/RBBGCMuso/RBBGCMuso")
```

> library("RBBGCMuso")

Ahogy a módszertan résznél már jeleztem, alapesetben a csomag bemeneti fájlokat nem állít elő a modell részére, azokat a felhasználónak kell előkészítenie. Néhány eset ez alól kivétel. Jelenleg több ún. REST-API (HTTP alapú alkalmazás-programozási interfész) is elérhető az Interneten, amelyeken keresztül hozzávetőleges (térben, időben interpolált) adatokat tudunk lekérni, majd ezekből előállítani a megfelelő, modell számára olvasható input fájt.

Kristóf Erzsébet hozzájárulása révén ma már lehetőség van arra, hogy az ERA5 reanalízis adatbázisból meteorológiai adatokat tötsünk le a modell részére, közvetlenül a Biome-BGCMuSo által elvárt formátumban. Ennek a megvalósítása még nem tökéletes, elsősorban az ERA5 kiszolgáló oldal lassúsága miatt, de hosszú távon szeretnénk ezt az eszközt hatékonnyá tenni. A jelenleg béta állapotú *getMeteo* függvény meteorológiai adatfájt állít elő a föld tetszőleges (földrajzi szélesség, illetve hosszúság alapján definiált) pontjára adott dőszakra vonatkozóan. Lehetőség van talajadatok lekérésére is, ami a Biome-BGCMuSo által elvárt talajfájl formátumában kér le adatokat a soilGrids adatbázisból (<u>https://soilgrids.org/</u>). Ennek megvalósítása a *createSoilFile* függvény által történik. A két lekérés parancsora nagyon egyszerű, és programozói tudás nélkül is használható:

```
> getMeteo(30,40)
```

> createSoilFile(30,40)

Az RBBGCMuso csomag kapcsán elkészítettünk egy kezdő modellező csomagot is, ami a *copyMusoExampleTo* paranccsal érhető el. Ennek segítségével a felhasználó kap egy teljes értékű szimulációs csomagot, beleértve az INI, EPC, meteorológia, talaj, CO₂, nitrogén ülepedés és menedzsment adatokat is a hegyhátsáli kísérleti állomás gyepterületére, kiegészítve az eddy kovariancia alapú mérési adatokkal (Barcza et al., 2003). A csomag azonnal futtatható, és az

eredmények elemezhetőek a magas szintű RBBGCMuso parancsokkal. A modellező csomag használatához ezt a parancsot kell kiadni:

> copyMusoExampleTo("hhs","C:\\muso example")

A jövőben más ökoszisztémákra paraméterezett példák közül is lehet majd választani, így segítve a felhasználót, hogy a saját modellezett területéhez leginkább hasonló beállításokból kiindulva tudja a modellt futtatni. Ezt követően a kutató az RBBGCMuso-val már minden beállítást, módosítást véghez vihet.

Összefoglalásként a 4. táblázatban az RBBGCMuso fontosabb függvényeit soroltam fel. További dokumentáció az RBBGCMuso csomag súgójában található a GitHub-on (https://github.com/hollorol/RBBGCMuso).

4. táblázat. A legfontosabb RBBGCMuso függvények listája.

Függvény neve	Funkcionalitás
setupMuso	A modell használatához szükséges központi adatstruktúrát állítja elő, infor-
	mációkat tartalmaz a felhasznált változókról, és a bemenő fájlokról. Kime-
	nete a "settings" objektum.
spinupMuso	Spin-up futást végez (a settings objektum alapján).
normalMuso	Lefuttatja a modellt normal módban (a settings objektum alapján)
runMuso	A spinupMuso és a normalMuso együttese, továbbá a két futás közti átme-
	netet is ellenőrzi. Ha a spinup nem fut le, akkor a normal sem. Opcionálisan
	ki tudja hagyni a spin-up függvényt. A kimenetet MS Excelbe/csv-be/txt-
	be, ods-be, NetCDF-be tudja exportálni
plotMuso	A modell lefuttatása után az eredmények ábrázolására nyújt megoldást.
	Amennyiben a modell még nem futott le, akkor ezt is elvégzi, képes a futta-
	tás előtt megváltoztatni az EPC vagy SOIL fájlokat, továbbá új kimeneti
	változók is definiálhatóak benne.
musoMonte	Hit and Run sampling algoritmust implementáló modul.
musoQuickEffect	Egy tetszőleges paramétert változtatva egy adott tartományban annak adott
	számú osztópontjában lefuttatja a modellt, és egy tetszőleges változó adott
	éves menetének alakulását vizualizáló eszköz
parameterSweep	A musoQuickEffect függvényt hatja végre tetszőleges számú paraméteren,
	majd a kimeneti grafikonokat egyetlen HTML dokumentumba exportálja,
	ahova a képelemek base64s kódolással kerülnek be, így a dokumentum füg-
	getlen lesz a képi hivatkozásoktól.
musoSensi	Sobol-érzékenységelemzést megvalósító függvény (Verbeeck et al., 2006)
updateMusoMap-	A kimeneti változókat és kódjaikat összekötő mátrix generálását végző
ping	függvény tetszőleges modellverzió output_map_init.c fájlát alapul véve.
musoMapping	Adott kimeneti változó kódjához tartozó változót megkereső függvény
musoMapping-	A kimeneti változók között tudunk ún. "fuzzy matching"-et végezni ezzel a
Find	függvénnyel, így könnyűszerrel megtalálhatjuk egy adott, számunkra érde-
	kes változó kódját.

Az alábbiakban a legalapvetőbb RBBGCMuso függvényeket mutatom be gyakorlati példákkal illusztrálva. A magas szintű GLUE, illetve CIRM alapú optimalizálást külön alfejezetben mutatom be, tekintve a dolgozatomban betöltött szerepét.

4.1.1. Ábrázolás, grafikon készítés

Az RBBGCMuso csomag grafikus feladatait a *plotMuso* függvény látja el, amely segítségével képesek vagyunk a modell futtatására, kiválasztott változók ábrázolására különböző aggregációs szinten.

A 18. ábrán látható példa a *plotMuso* függvény egyszerűségét hivatott bemutatni. Csupán egy öt soros R kód segítségével lefuttattam Martonvásárra a modellt és ábrázoltam az eredményeket a 2010-es évre.



18. ábra. plotMuso példa. Kukorica szimuláció eredménye Martonvásárra.

A szimulációk kukoricára vonatkoznak, és a bemutatott martonvásári esettanulmány alapján készültek. Ez magyarázza a napi fotoszintézis (GPP-ben kifejezett értkékének) alakulását az éven belül, ami egy relatíve rövid vegetációs időszakot jelez.

4.1.2. Lokális érzékenység-elemzés

Minthogy a lokális érzékenység-elemzés legfőbb célja az adott paraméter viselkedésének tanulmányozása a többi érték állandóan tartása mellett, a csomagban kizárólag kvalitatív, grafikus megoldást kínálok. Ez a rutin a *musoQuickEffect*. A függvény egy adott – *startVal* és *endVal* által meghatározott – intervallumon *nSteps* számú osztóponton futtatja a modellt, majd ábrázolja azt egy adott kimeneti változóra nézve. A függvényben kiválasztott évre is lehet szűkíteni. Használatát a következő 11 sor demonstrálja, amelynek eredményét a 19. ábrán lehet látni.





19. ábra. musoQuickEffect: adott paraméter értékének hatásának egy adott változóra. Ez a példa a bázishőmérséklet (növénynövekedést érintő hőmérsékleti küszöb) megválasztásának hatását mutatja.

4.1.3. Globális érzékenység-elemzés

A lokális érzékenység-elemzéshez hasonlóan az RBBGCMuso csomag megoldást kínál globális érzékenység-elemzésre is. A Sobol-érzékenység alapú relatív érzékenységértékek

kiszámítását a *musoSensi* függvény segítségével lehet elvégezni. (lásd. a 3.2.1. fejezet konvenciókra vonatkozó részt). Az elemzés elkészítéséhez a korábban említett *parameters.csv* fájl elkészítésére van szükség, amelyben felsoroljuk a vizsgálandó paramétereket, illetve azok minimum és maximum értékét. Az előkészítés után egy egyszerű sorral elvégezhető a vizsgálat:

> musoSensi()

A parancs kiadása után a nagy számú, párhuzamosan futtatott randomizáció után készül el az eredmény ábra, ami vizuálisan értelmezhető (erre egy példa a 20. ábra).



20. ábra. Példa érzékenység-elemzés eredményére az RBBGCMuso csomag implementációja alapján. A vízszintes tengelyen az ökofiziológiai bemeneti fájl néhány kiválasztott parametére olvasható, míg a relatív érzékenység az y tengelyen látható. Az eredmények alapján a Rubisco enzimben tárlolt nitrogént (FLNR) meghatározó paraméter és a kelés időpontjának fontossága nyilvánvaló.

Az RBBGCMuso csomaggal GLUE, illetve Bayes alapú inverz paraméter-becslés is végrehajtható (lásd <u>https://github.com/hollorol/RBBGCMuso</u>), amit a következő alfejezetben az általam kifejlesztett új CIRM módszerrel együtt mutatok be.

4.2. GLUE és CIRM

4.2.1. GLUE és utófeldolgozás a döntési fákon keresztül

A 3.6.1.5 alfejezetben említettek alapján az új optimalizációs eljárást GLUE környezetben valósítottam meg az RBBGCMuso szoftverrel, iteratív módon. Az iteráció 10 lépésből áll, és az algoritmus szerint az egyes iterációs lépések után a következő mindig a módosított partaméterintervallumokat használja kiindulásként.

A 21. ábra az első iterációs lépés jellegzetes GLUE alapú pontdiagramjait mutatja. A 20 vizsgált paraméterből (2. táblázat) kiválasztottunk 4-et, amelyek az utolsó (10.) iterációs lépés (lásd alább) tipikus, minden esetre jellemző mintázatát képviselik.



21. ábra. A CIRM alapú optimalizálási eljárás kiválasztott pontdiagramjai az első iterációs lépés után (az összes szimulációs eredményt ábrázoltam, beleértve a jól viselkedő és nem jól viselkedő futtatásokat). A szürke pontok az összes szimuláció valószínűségi értékeit (marginális eloszlásokat), míg a piros pontok az elfogadható szimulációkhoz tartozó valószínűségi értékeket mutatják. A fekete függőleges vonal a ML becslésen alapuló paraméterértéket jelöli, míg a kék függőleges vonal az elfogadható szimulációk mediánértékét mutatja az első iterációs lépés után.

Az ábrák leginkább szembetűnő jellemzője az ekvifinalitás az összes szimulációt ábrázoló szürke pontok alapján. Ez szinte minden más paraméterre is igaz, amit itt nem mutatunk be (Hollós et al, 2022, Supplement). Van néhány kivétel, ahol valamilyen mintázat felismerhető a grafikonokon

(maxlifetime4, leaf_allocation_3, leaf_allocation_4, stem_allocation_3 és bizonyos mértékig a Rubisco), de a szürke pontok eloszlása csak csekély mértékű intervallumszűkítést mutat (pl. stem_allocation_3, leaf_allocation_4 esetében). Más paraméterek esetében, mint például a maxlifetime4, még ha a pöttyös ábrákon fel is ismerhető valamilyen mintázat, a jól viselkedő szürke pontokon alapuló utólagos intervallumok nem eredményeznek paraméterintervallum-csökkentést. Nagyon fontos megjegyezni, hogy egy tipikus Bayes vagy GLUE alapú kalibrációban ez az optimalizálás utolsó szakasza, amely egyértelműen nem kielégítő és tulajdonképpen sikertelen az intervallum-csökkentésre való törekvés szempontjából.

Ezen a ponton kezdődik valójában a CIRM alkalmazása. Miután úgy döntöttünk, hogy az eredmények elfogadhatóságát az előre meghatározott szűrés alapján ellenőrizzük (a HI, a LAI_{max}, a virágzás időpontja és a gyökérmélység alapján), az összkép megváltozik (21. ábra; piros pontok). Azonban valószínűleg az optimalizálás nagy szabadsági foka miatt az első iteráció után csak 596 szimuláció volt megvalósítható, ami egyértelműen alacsony sikerességi arány (c_r= 5,96%). Az első iterációs lépést követően a "dotty-plot"-ok még mindig ekvifinalitást mutatnak az esetek többségében. Van néhány kivétel, mint például a Rubisco, maxlifetime4, leaf_allocation_3, root_allocation_3, stem_allocation_3, leaf_allocation_4, root_allocation_4, fruit_allocation_6 és stem_allocation_6, amelyek esetében a megkötés alapján végzett futtatás szerint szűkebb tartományoknak kellene lenniük (2. ábra; Hollós et al., 2022 Supplement). A Rubisco esetében például a piros pontok alapján a ~ [0,7 - 0,11] paraméter-tartomány tűnik észszerűnek. Itt ismét hangsúlyozni szeretném, hogy egy tipikus optimalizálási feladat során a felhasználó jellemzően csak a szürke pontok alapján vonja le a következtetéseket. Ebben az értelemben az elfogadható szimulációkból kinyert információk már egy lépéssel előrébb járnak, tehát már az új módszer által hozzáadott értéket vizsgáljuk.

Felismerhető, hogy a kis *c*_{*r*} miatt nem rendelkezünk elegendő számú elfogadható szimulációval ahhoz, hogy megbízzunk a GLUE optimális értékek, illetve a GLUE bizonytalansági tartományok jóságában. A *c*_{*r*} növelése érdekében a paraméter-intervallumok további finomítása az észszerű következő lépés. A felépített döntési fák információt nyújtanak a paraméterek és az elfogadható/nem elfogadható szimulációk közötti lehetséges kapcsolatokról, így hasznosak lehetnek a paraméterintervallumok frissítéséhez. Tekintettel arra, hogy 4 kimeneti megkötésünk van, 4 döntési fát készítettünk.

A 22. ábra az első feltétel (HI) alapján az 1. iterációs lépés után felépített döntési fát mutatja. A 2. rétegben a length_of_phenophase_4 paraméter vágási küszöbértéke 372. Ha a paraméter kisebb, mint 372, akkor a többi szinttel elérhetjük a megfelelő ágat. A fa alján helyezkednek el a levélcsomópontok. A levélcsomópontok azok homogenitása miatt az osztályozás eredményeit képviselik. A 22. ábrán a kék levélcsomópontok (azaz a 0-val jelölt lekerekített négyzetek az ábra alján, amelyek az $f \circ M$ eredményei) a nem elfogadható szimulációkat jelölik, míg a zöld levélcsomópontok az elfogadható szimulációkat (1-gyel jelölve). A levélcsomópontokon belüli százalékos értékek az adott ághoz tartozó szimulációk hányadát mutatják.



22. ábra. Döntési fa az első iterációs lépés után, a harvest indexre vonatkozó megkötés alapján. A levélcsomópontokon belüli százalékos értékek összege a kerekítés miatt nem 100%. A zöld szín a megvalósítható szimulációkat jelenti, a kék pedig a megvalósíthatatlan szimulációkat. A százalékos értékek azt mutatják, hogy a szimulációk hányad része tartozik az adott levélhez.

A 22. ábra szerint az alkalmazott HI-szűrő esetén a mintavételezett paraméterkombinációk 58%-a nem volt megvalósítható (a kék dobozokban szereplő százalékos arányok összege). Amint azt az algoritmus leírásában kifejtettem, a CIRM megközelítésünkben mindig arra a levélcsomópontra összpontosítok, amelyikben a megvalósítható szimulációk aránya a legmagasabb. Ebben az értelemben a döntési fa azt sugallja, hogy a HI-szűrőhöz kapcsolódó legfontosabb paraméterek csökkenő stem allocation 6, sorrendben а következők: length_of_phenophase_4, length_of_phenophase_3 és maxlitefime_4 (ez az út a felsőtől a legmagasabb %-kal rendelkező levélcsomópontig, miután kizártuk a paraméterek ismételt előfordulását a döntési faalsóbb rétegeiben). Ezek a paraméterek közvetve befolyásolják a modellben a szemtelítődést, és így irreális HI-t eredményeznek. Kissé meglepő, hogy a fruit_allocation_6 paraméter nem szerepel a döntési fában ezen az úton (de szerepel a többi levélcsomóponthoz vezető más úton). A Maxlifetime4 befolyásolja a levelek leszáradásának dinamikáját a szemtelítődés előtt és alatt, így közvetlen kapcsolatban áll HI-vel. A 3. és 4. fenofázis hossza befolvásolja a levélnövekedés dinamikáját, amely egyértelműen befolyásolja az asszimilációt, így a szemtelítődést a 6. fenofázisban. A a döntési fából nyert információk alapvetően hasznosak, és betekintést nyújtanak a Biome-BGCMuSo-ban megvalósított növényi növekedés és a végső termésmennyiség kialakulásának összetett folyamatába.

A legnagyobb sikerességi arányhoz (21%, a jobb szélső levélcsomópont) tartozó levélcsomóponthoz vezető útvonal mentén hozott döntések alapján kiszámíthatjuk a küszöbértékeket, és frissíthetjük az eredeti paraméterintervallumokat (lásd a 2. táblázatot). A szár allokáció a 6. fenofázisban 0,31-nél kisebbnek és 0,16-nál nagyobbnak kell lennie (az eredeti intervallum [0,1-0,4] volt; lásd a 2. táblázatot). A 3. fenofázis hosszának kisebbnek kell lennie 405-nél, a 4. fenofázis hosszának pedig kisebbnek kell lennie 372-nél. A Maxlifetime4 értéket 693-nál nagyobbra kell beállítani.

Megjegyzendő, hogy az eljárás bármelyik szakaszában a felhasználó választhat egy másik levélcsomópontot alacsonyabb sikerességi aránnyal, ha a szakértői tudás egy alternatív választást támogat. Ebben az esetben szükség lehet a paramétertartományok kézi módosítására, és az iterációt a módosított intervallumokkal kell újraindítani (a fent leírt parameters.csv alapján).

A 23. ábra a a LAI_{max} feltételhez tartozó döntési fát ábrázolja az 1. iterációs lépés után. Az ábrán látható, hogy a mintavételezett paraméterkombinációk 58%-a nem elfogadható.



23. ábra. Döntési fa az első iterációs lépés után a LAI_{max} feltétel alapján. Megjegyzendő, hogy a levélcsomópontokon belüli százalékos értékek összege a kerekítés miatt nem 100%. A zöld szín a megvalósítható szimulációkat jelenti, a kék pedig a megvalósíthatatlan szimulációkat. A százalékos értékek azt mutatják, hogy a szimulációk hányad része tartozik az adott levélhez.

A fa alapján a legfontosabb paraméterek (csökkenő sorrendben) a Rubisco, a leaf_allocation_4 és a length_of_phenophase3. Az alsóbb rétegekben a Rubisco és a leaf_allocation_4 ismét megjelenik. Mivel a Rubisco végső soron a maximális fotoszintézis mértékét szabályozza (White et al., 2000), fontossága a levélfejlődés szempontjából egyértelmű. A levélallokáció szerepe a 4. fenofázisban (amely meghatározza a maximum LAI-t) szintén egyértelmű és könnyen értelmezhető.

A 3. fenofázis hossza kevésbé intuitív, de ésszerű, mivel befolyásolja a levélfejlődés kezdeti állapotát a 4. fenofázisban, amikor a LAI eléri a maximumát.

A döntési fa jobb szélső levélcsomóponthoz vezető út (amelyhez a legnagyobb, 16%-os sikerességi arány társul) segítségével új intervallumokat állíthatunk be a paraméterekhez. A 23. ábrán látható döntési fa összes döntési csomópontja alapján a Rubisco értékének kisebbnek kell lennie 0,097-nél és nagyobbnak 0,079-nél, a leaf_allocation_4 értékének kisebbnek kell lennie 0,34-nél, a length_of_phenophase3 értékének pedig nagyobbnak kell lennie 296-nál. Megjegyzendő, hogy a 22. ábrán bemutatott döntési fa már meghatározta a length_of_phenophase3 új felső határért, amelyet itt tovább finomítunk.

A 24. ábra mutatja be az 1. iterációs lépésre vonatkozóan a gyökérmélység-megkötés alapján konstruált döntési fát. Ebben az esetben a mintavételezett paraméterkombinációk mindössze 29%-a volt nem elfogadható. A fa azt sugallja, hogy a legfontosabb kapcsolódó paraméterek a Rubisco és a length_of_phenophase_3. Ez ésszerű, figyelembe véve a Rubisco meghatározó szerepét az általános produkció szempontjából, és figyelembe véve a 3. fenofázis fontosságát a gyökérfejlődés szempontjából.



24. ábra. Döntési fa az első iterációs lépés után a gyökérmélység-megkötések alapján. A levélcsomópontokon belüli százalékos értékek összege a kerekítés miatt nem 100% A zöld szín a megvalósítható szimulációkat jelenti, a kék pedig a megvalósíthatatlan szimulációkat. A százalékos értékek azt mutatják, hogy a szimulációk hányad része tartozik az adott levélhez.

A Rubisco intervallum itt tovább finomítható (0,09-nél nagyobbnak kell lennie, és a 23. ábrán látható, előző fa szerint 0,097-nél kisebbnek kell lennie). A Length_of_phenophase3-nak nagyobbnak kell lennie 294-nél, amit ebben a szakaszban valójában nem használunk, mivel már 296-nál nagyobbnak volt beállítva (22. ábra).

A 25. ábrán a virágzási időn alapuló szűrőfüggvény segítségével készült döntési fát mutatja be. A virágzás időpontjára vonatkozó döntési fa használatával a mintavételezett paraméterkombinációk 42%-a megvalósíthatatlannak bizonyult.



25. ábra. Az első iterációs lépést ábrázoló döntési fa virágzás időpontjára vonatkozó megkötés alapján. A levélcsomópontokon belüli százalékos értékek összege a kerekítés miatt nem 100%. A zöld szín a megvalósítható szimulációkat jelenti, a kék pedig a megvalósíthatatlan szimulációkat. A százalékos értékek azt mutatják, hogy a szimulációk hányad része tartozik az adott levélhez.

A fa azt sugallja, hogy a legfontosabb paraméterek a 3. és 4. fenofázis hossza. Ez tökéletesen megfelel a várakozásoknak, mivel a virágzás időpontját az azt megelőző fenofázisok hossza határozza meg GDD-ben kifejezve (megjegyzendő, hogy az első két fenofázis hossza ezen kalibráció során rögzített).

A döntési fa iránymutatást ad a két paraméter frissítéséhez. A jobb szélső levélcsomóponthoz vezető út szerint a length_of_phenophase3 értékét 318-nál nagyobbra, a length_of_phenophase4 értékét pedig 313-nál nagyobbra kell beállítani. Ezek az új beállítások tovább korlátozzák ezeket a paramétereket, mivel azok már bizonyos mértékig finomítva voltak.

Ebben a szakaszban a legfontosabb felismerés a döntési fák hasznossága, amelyek a paraméter-intervallumok kézi frissítésére használhatók. Ez a fajta információ eddig "rejtve" volt, mivel a marginális eloszlások nem mutatták a paraméter-intervallum csökkentésének lehetőségeit. Az egyes szűrőfüggvényekre vonatkozó döntési fák elemzések elvégzése után az egyes fákból származó feltételeket kombináljuk, és módosítjuk a paraméter-intervallumokat.

4.2.2. Az iterációk eredményei és a döntési fák automatikus értelmezése

Az előző pontban leszűkített paraméterintervallumok a módszer megismétlésével tovább szűkíthetőek (ld. 1. algoritmus).



26. ábra. A baloldali ábrán a sikerességi arányt láthatjuk az iterációs lépések számának függvényében az automatizált munkafolyamat során, míg a jobboldali ábrán a szűrőfüggvényeinkre vonatkozó döntési fák pontosságának változását láthatjuk az iterációs szám függvényében. (A rövidítések a 2. táblázatban találhatóak)

A szűkítések iteratív elvégzése során a sikerességi arány (26. ábra) monoton nőtt, ami jelzi, hogy a szűkített intervallumokban a megbízható szimulációk száma növekedett. Láthatjuk, hogy a white-box közelítés is helyes volt, mert átlagosan a közelítés pontossága monoton nőtt, habár a pontosság metrikája és maga az osztályozó döntési-fa algoritmus is akkor megbízható, ha a kategóriák hasonló arányban lelhetők fel. A sikerességi arány és a pontosság együtt történő növekedése viszont jó indikáció arra vonatkozóan, hogy a módszer hatékonyan működött.

A 27. ábra az utolsó lépésből kiválasztott, tipikus mintázatokat reprezentáló pontdiagramokat mutatja. A grafikonon látható, hogy ellentétben a 21. ábrával (1. lépés), szinte minden szimuláció elfogadható volt ebben a szakaszban (a szürke pontok alig észlelhetők). Bár a legtöbb paramétert ekvifinalitás jellemzi, a paraméter-intervallumok lényegesen kisebbek, mint az első lépés után. Néhány paraméter jól kimutatható optimumot ad (pl. max_lifetime_4 a 27. ábrán), tipikus paramétereloszlással. Egyes paraméterek esetében – mint például a stem_allocation_4 esetén – jól körülhatárolható pontfelhőt láthatunk az ábrán. A 3. és 4. fenofázis hosszát szabályozó paraméterek szokatlan eloszlást mutatnak, ami egyértelműen a döntési fák által meghatározott határértékek következménye (lásd fent). Ezeken az ábrákon a jól viselkedő paraméterek lokalizálják az optimális értéket.


27. ábra. A 10. iterációs lépés utáni optimalizálás kiválasztott pontdiagramjai. A szimbólumok és a függőleges vonalak jelentése megegyezik a 21. ábrával

A 10. iterációs lépés után majdnem minden mintavételezett paraméter elfogadható volt (95,45%, azaz 10 000-ből 9545 iteráció volt elfogadható). Ez azt jelenti, hogy majdnem minden szimuláció megfelelt az előre meghatározott megkötéseknek, és így a felhasználó elvárásainak megfelelő eredményeket adott. Tekintettel a sikeres és értelmes szimulációk nagy számára, a magas mintaszám egyértelműen javítja a felhasználó bizalmát az eredmények statisztikai tulajdonságai (leginkább az optimum és a bizonytalansági tartományok) tekintetében.

A paraméterek optimális értékeit figyelembe véve (amik a pont-diagrammokon ábrázolt függőleges vonalak) a maximális valószínűségi értékek jellemzően eltértek a "jól viselkedő" adatokból számított értékektől (azaz a GLUE mediánjától), hasonlóan az 1. lépéshez (Hollós et al., 2022, Supplement).

A feltételes tartománycsökkentő eljárás lépéseit a ML értékekkel együtt a 28. ábra mutatja be. Az ábra segít értelmezni az iteratív intervallum szűkítések hatását az ML értékekre illetve a paraméter-bizonytalansági tartományra nézve. Megjegyezzük, hogy a 10. iterációs lépésnél a GLUEalapú intervallum-csökkentést is figyelembe vettük. A GLUE-alapú optimum nincs feltüntetve, de ez a tényleges intervallum középpontjával közelíthető.

A 28. ábra egyértelműen mutatja, hogy a paraméter-intervallumok sok esetben jelentősen csökkentek. Egyes paraméterek, például a Rubisco és a length_of_phenophase4 esetében az

intervallum csökkenése az első iterációs lépésben volt a legnagyobb. Más paraméterek, például a root_depth_function_shape, leaf_allocation_3 és stem_allocation_6 esetében a csökkenés fokozatos volt, és nem feltétlenül egy iterációs lépéshez kapcsolódott. Ez a többlépéses megközelítés hasznosságát jelzi. Néhány paraméter esetében az a priori intervallum változatlan maradt (pl. root_allocation_3, stem_allocation_3). A grafikon azt is mutatja, hogy néhány esetben az ML-érték kívül esett a végső paraméterintervallumon (azaz a 10. lépésben), ami nem elfogadható megoldást jelez. Mivel a tradicionális kalibrációs eljárásoknál, mint amilyen a Bayes-alapú vagy a frekventista, már az első iterációs lépésnél megállunk, az eredmények egyértelműen mutatják a CIRM előnyét, ami révén elkerülhető, hogy "jó eredményt rossz okokból" kapjunk.



28. ábra. A bemutatott új módszer teljesítményét bemutató ábrák. A vonalak egy adott paraméter alsó és felső határértékét és azok változását jelzik az egymást követő iterációs lépések között, míg a körök az adott iterációs lépésre vonatkozó ML becslési eredményeket jelzik. A rövidítések jelentését a 2. táblázat tartalmazza.

Az 5. táblázat összefoglalja a modell-optimalizálás eredményeit, információt nyújt a poszterior paraméter-intervallumokról, a végső lépésben az ML becslést jelentő paraméterkészletről és a GLUE-alapú paraméter-intervallumokról. Az intervallum hosszának átlagos változása 44% volt a 20 vizsgált paraméter esetében (a legnagyobb érték 88% volt a Rubisco esetén).

Nem történt paraméter-tartomány csökkenés a length_of_phenophase_6, a root_allocation_3, a stem_allocation_3, a root_allocation_6 és a maxlifetime_3 esetében. A szár- és gyökérelosztás a 3. fenofázisban kevésbé tűnik meghatározónak a végső terméshozam szempontjából, ami érdekes eredmény. A 6. fenofázisból származó paraméterek viselkedésének értelmezése egyszerűbb. E fenofázis pontos hosszának becslése lehetetlen, mivel a levelek leszáradása után a szemtelítődés során a végső terméshozam már nem változhat, így értékét semmilyen megfigyeléssel vagy kényszerrel nem lehet meghatározni. A gyökérelosztás a 6. fenofázisban kicsi, és úgy tűnik, hogy nincs jelentős hatása a végső terméshozamra.

5. táblázat. Az optimalizált paraméterek listája 10 iteráció után, a megkötésekkel együtt. Az ML és a
GLUE-alapú optimalizált paraméterértékek is fel vannak tüntetve. Az intervallum hosszának
százalékos változása az utolsó (10.) iterációs lépést követően az apriori intervallum hosszához
képest (vö. 2. táblázat).

Rövidítés	MIN	MAX	ML	GLUE	Százalékos
					változás
Rubisco	0,09029	0,09615	0,0903	0,0929	88
root_distribution_paraméter	2	4,966	4,8722	3,629	26
root_weight_to_max_root_depth	0,08	0,1053	0,09973	0,0920	58
root_depth_function_shape	0,4	1,118	0,7649	0,7583	40
senescence_coeff_for_leaf	0,001	0,03323	0,00255	0,01469	17
vízstressz hatása a fotoszintézisre	0	0,4	0,3654	0,1886	43
length_of_phenophase_3	334,2	367,4	345,96	342,63	84
length_of_phenophase_4	313,9	362,8	314,279	320,392	77
length_of_phenophase_6	851,3	1200	968,46	1007,82	0
leaf_allocation_3	0,4	0,4412	0,4096	0,417	59
root_allocation_3	0,3	0,5	0,3429	0,339	0
stem_allocation_3	0,2	0,5	0,2473	0,244	0
leaf_allocation_4	0,2537	0,2932	0,2709	0,273	87
root_allocation_4	0,2689	0,4	0,3987	0,334	34
stem_allocation_4	0,306	0,4841	0,3302	0,393	11
root_allocation_6	0,05	0,15	0,1234	0,095	0
fruit_allocation_6	0,6299	0,6933	0,64670	0,657	79
stem_allocation_6	0,2149	0,2824	0,22983	0,248	78
maxlifetime_3	500	1300	540,46	930,70	0
maxlifetime_4	866,8	1300	1018,27	965,83	46

Azon paraméterek esetében, ahol 0%-os volt a paraméter-intervallum csökkenés, további megkötések alkalmazása szükséges. Megjegyzendő, hogy bizonyos esetekben ez nem jelent

problémát, mivel a többi függő paraméter már meghatározza e paraméterek értékét (mint például az allokáció esetében, amikor a paraméterek összegének 1-re kell összeadódnia).

Összességében a szűkített paraméter-intervallumok használata 42,3%-os csökkenést eredményezett a szimulált hozam bizonytalanságában, amelyet a modellezési eredmények évente kiszámított standard eltérésének átlaga alapján számoltam ki, az eredeti és a csökkentett paraméter-tartományokból Monte Carlo alapú mintavételezéssel végzett 1000 szimuláció alapján (29. ábra)



29. ábra. A modellezett termésmennyiség bizonytalansága a prior (felső ábra) és poszterior (alsó ábra) paraméterezéssel. A zöld háromszögek a megfigyeléseket mutatják, míg a fehér körök a szimulációk mediánját reprezentálják. 1000 mintavétel alapján történt az eredmények származtatása.

4.2.3. A modell teljesítményének számszerűsítése

A 30. ábra a modell eredményeit mutatja be az a priori paraméterezés és a GLUE-alapú, optimalizált paraméterkészlet esetén. Az ábrán látható, hogy a megfigyelések bizonytalansága nagy, mivel a martonvásári kukorica- adatok több kísérlet parcelláinak összevonása által lettek származtatva (3.3 fejezet). Az ábra azt mutatja, hogy az optimalizált modell sok esetben a bizonytalansági tartományon belül becsülte a terméshozamot.

A 6. táblázat a szimulációk statisztikai értékelésének eredményeit mutatja. Itt szerepelnek az ML értékkel rendelkező szimulációk teljesítménymutatói is. A táblázat a szimulációk minőségének

jelentős javulását mutatja az egymást követő iterációs lépésekben. A magyarázott variancia jellemzően magasabb volt a ML módszer, mint a GLUE esetében, az előbbinél az R² már az 1. iteráció után jelentősen nőtt, míg az utóbbi esetében ez csak a 6. iteráció után következett be. Továbbá az ML módszer esetén az RMSE értékek alacsonyabbnak bizonyultak a GLUE-hoz viszonyítva (kivétel a 6. lépést). A szisztematikus hiba (bias) változó volt, de ML esetében 0 közelébe került, míg a GLUE esetében pozitív maradt. Az ME értékek alapján a kalibrációs adathalmazon a ML módszer jobbnak bizonyult, mint a GLUE. Összefoglalva, minden mutató tekintetében jobb teljesítményt látunk a ML paraméterezés esetében, ami túlillesztésre utalhat.



30. ábra. A megfigyelt átlagos kukoricatermés (teli körök; a bizonytalanság +/- egy szórásként van feltüntetve), az a priori szimulált (üres háromszögek) és az optimalizált (GLUE-alapú; zöld háromszögek) kukoricatermés idősorai Martonvásáron 1991-től 2018-ig.

Paraméter-									
becslés	ML				GLUE				
	R ²	RMSE	bias	ME	R ²	RMSE	bias	ME	
a priori	0,16	4,752	-4,448	-5,793	0,16	4,752	-4,448	-5,793	
1. lépés	0,37	1,548	0,486	0,279	0,07	2,209	0,324	-0,468	
2. lépés	0,41	1,411	-0,003	0,401	0,28	1,675	0,051	0,156	
3. lépés	0,51	1,326	0,176	0,471	0,02	2,408	-0,063	-0,744	
4. lépés	0,44	1,458	0,449	0,360	0,33	1,689	0,268	0,142	
5. lépés	0,45	1,388	0,043	0,421	0,33	1,678	0,158	0,153	
6. lépés	0,36	1,552	0,478	0,275	0,41	1,552	0,269	0,276	
7. lépés	0,41	1,461	0,167	0,358	0,38	1,573	0,377	0,256	
8. lépés	0,46	1,391	0,002	0,418	0,38	1,548	0,270	0,280	
9. lépés	0,40	1,465	-0,045	0,354	0,39	1,558	0,292	0,270	
10. lépés	0,48	1,358	-0,102	0,446	0,38	1,585	0,395	0,244	

6. táblázat. A modell teljesítményének statisztikai értékelése az egyes iterációs lépések után a martonvásári kukoricatermés hosszú távú (1991-2018) szántóföldi kísérleti adatainak felhasználásával.

A 7. táblázat a kukoricatermésre vonatkozó modellkimenet teljesítménymutatóit mutatja Martonvásáron és a független validálási kísérlet (megyei szintű szimuláció) esetében. Mind a kalibrációs halmaz (Martonvásár), mind a validációs halmaz esetében az optimalizált szimulációs eredmények jelentősen közelebb voltak a megfigyelésekhez, mint az a priori szimulációk. Ez azt jelenti, hogy a kalibrált modell jobban alkalmazható a megyei szintű független adathalmazra.

A 10. lépésből származó ML paraméterezés a bias és az ME tekintetében felülmúlta a GLUEalapú szimulációkat, míg az RMSE-értékek közötti különbség kicsi volt. Az R² tekintetében a GLUEalapú módszer jobban teljesített.

A magyarázott variancia (48%) itt valójában magasabb volt, mint az ML esetében. Megjegyzendő, hogy az ME mindkét esetben negatív volt, mivel a termés nagysága Martonvásáron és megyei szinten eltérő volt, valószínűleg az eltérő műtrágyázási szint és agrotechnika miatt.

7. táblázat. A kukoricatermés különböző szimulációinak hibastatisztikái. Martonvásár a kalibrációs adathalmazt, míg Fejér megye a független, KSH adatokon alapuló modellértékelést jelenti a különböző paraméterezésekkel. Fejér megye esetében a végleges, optimalizált paraméterkészletet használtuk (5. táblázat). A részleteket lásd a szövegben.

	R ²	RMSE	bias	ME
a priori Martonvásáron	0,16	4,752	-4,45	-5,793
a priori Fejér megyében	0,15	3,500	-3,08	-2,658
10. lépés Martonvásáron (ML)	0,48	1,358	-0,10	0,446
10. lépés Martonvásáron (GLUE)	0,38	1,585	0,39	0,244
10. lépés Fejér megyében (ML)	0,37	1,986	1,28	-0,109
10. lépés Fejér megyében (GLUE)	0,48	2,028	1,49	-0,157

A GLUE jobb teljesítménye a validációs kísérletben az R² tekintetében azt jelzi, hogy a ML túlillesztéssel járt, és a GLUE-alapú módszer talán jobban alkalmazható. Megjegyzendő, hogy ebben a szakaszban a modellező hibrid paraméterezést végezhet, amely egyes paraméterek esetében, amelyek kicsi vagy nulla közeli intervallumcsökkenéssel járnak, a ML értékeket, más esetekben pedig GLUE-alapú, korlátozottabb paramétereket használ. Egyes paraméterek ML értékei informatívak lehetnek, ha a több iterációs lépés során konszenzus van az értékeikben.

Megjegyzendő, hogy az alacsony magyarázott variancia nem feltétlenül jelent rossz szimulációt. Esetünkben a likelihood normális volt, így a optimalizációs eljárás során a fő cél a hiba minimalizálása volt, nem pedig az R² maximalizálása.

4.3. Az AgroMo keretrendszer

Az előző fejezetekben ismertetett RBBGCMuso csomag, valamint a kalibrációs megoldások kiterjesztése és a funkcionalitások felhasználóbaráttá tétele során létrejött az AgroMo rendszer (amelynek kezdőképernyője a 31. ábrán látható).



31. ábra. AgroMo nyitólap.

Az AgroMo egy olyan keretrendszer, ami integrálja a klímaadatokat, a talajadatokat, a biogeokémiai modellt a különböző mezőgazdasági döntések forgatókönyveivel. A rendszer, nemzetközi együttműködések keretében 9 nyelven érhető el, s a felületét több, a fejlesztésben nem érintett külföldi felhasználó is értékelte, illetve alkalmazásának lehetőségeit kereste. Erre bizonyíték, hogy a rendszert továbbfejlesztés céljából a GitHub-on 5-en "forkolták" (2023.10.31-ig), ami annyit jelent, hogy egy saját fejlesztési ágat ágaztattak le a fő ágról. A rendszert továbbá 11 fejlesztő értékelte pozitívan, valamint beválogatták az "Open Sustainable Technology" (https://github.com/protontypes/open-sustainable-technology) lektorált nemzetközi válogatásba.

Az AgroMo 5 + 1 fő részből áll (Site, Grid+Query, Plot, MAP, ParAna). A továbbiakban az elkészült rendszert ezek részletes leírásával fogom bemutatni.

4.3.1. AgroMo Site (cella-szintű futtatások)

Mivel az AgroMo központi modellje (Biome-BGCMuSo) pontszintű (horizontális heterogenitást nem veszi figyelembe), és a térbeli szimulációkat cella-szintű szimulációk együttesével közelítjük, a rácshálóra történő kiterjesztés nagymértékben függ a pontszintű működéstől. Ebből a célból alkottuk az AgroMo Site felületet (32. ábra), ahol cella-szintű futtatásokat tudunk végezni hibakeresés, vagy önálló futtatás céljából.

	INI file:		CELL id	:		_
	Martonvasar_i	naize.ini	• 1	•	SITE	
	WEATHER file:					
-	Martonvasar.w	rth			•	
site	SOIL file:					
	Martonvasar.s	oi			•	
PLOT GRID	MANAGEMENT fi	le:				
BASE MAP	maize.mgm				•	
	management op	ntions:	change (+/-):			
	planting:	maize.plt	date (day):	0	density (p/m²):	0
0	harvest:	maize.hrv	date (day):	0		
	fertilization:	maize.frz	date (day):	0	amount (kg/ha):	0
0	irrigation:		date (day):	0	amount (mm):	0
	cultivation:		•			
	grazing:		•			
	mowing:		•			
<u>o</u>	thinning:		•			
	OUTPUT DATA T	ABLE:				
O	Martonvasar_I	maize				
В		START SIMULATION		Ρ	LOT	

32. ábra. AgroMo Site.

A pontszintű szántóföldi modellezés fontossága a mezőgazdasági kutatásban és döntéshozatalban nem hagyható figyelmen kívül. Az ilyen típusú modellek lehetővé teszik számunkra, hogy aprólékosan megvizsgáljuk és megértsük, hogyan reagál egy adott kultúra különféle változó környezeti és gazdálkodási feltételekre. Ezen belül, különösen a vetés időzítését és a különböző agrotechnikai műveletek – mint például az öntözés, trágyázás vagy növényvédelem – hatásait tudjuk modellezni és elemezni. Az ilyen cellaszintű modellezéssel szerezhető adatok és ismeretek nélkülözhetetlenek a mezőgazdasági döntések meghozatalában, mivel az alapvető intuíciónkat formálják a növényállomány reakcióiról, és ezáltal a gazdálkodók jobb, megalapozottabb döntéseket hozhatnak.

Emellett a pontszintű modellezés eredményei a kutatások széleskörűbb kiterjesztésének alapját is képezik. A megértett és jól modellezett lokális viselkedési minták lehetővé teszik a kutatók számára, hogy megfogalmazzák és validálják azokat a hipotéziseket, amelyek a rácshálós modellekben vagy térbeli kiterjesztésekben használhatók. Ezek a térbeli modellek aztán a mezőgazdasági területek nagyobb skálájú elemzését teszik lehetővé. Ennek eredményeképpen a pontszintű modellezés nem csupán a mezőgazdasági gyakorlatban való közvetlen alkalmazásra szolgál, hanem híd szerepét is betölti a mikro és makro szintű kutatások között. Az AgroMo Site platform kifejezetten ezen pontszintű szántóföldi modellezési futtatások támogatására lett kifejlesztve, és olyan eszközöket nyújt, melyekkel a felhasználók testre szabhatják és optimalizálhatják a modellezési környezetet. A platform lehetőséget biztosít a megfelelő beállító fájlok (INI fájl) kiválasztására, a meteorológiai adatsorok behívására és az ökofiziológiai paraméterek beállítására az AgroMo rendszerén belüli mappákban található adatfájlok segítségével (erről a mapparendszerről későbbiekben részletesebben írok a 3.7.3 fejezetben.). Továbbá, a felhasználók a management fájlokat is kiválaszthatják a rendelkezésre álló adatbázisból. Amikor többféle beállító fájl áll rendelkezésre, a felhasználók előre definiált kombinációkat alkalmazhatnak, vagy saját kombinációkat hozhatnak létre, hogy új és új kérdésekre keressék a választ a modellezés során.

A platform további rugalmasságot nyújt a management döntések szimulációjában. Lehetőség van például a vetés és aratás időpontjának relatív beállítására vagy módosítására, illetve az öntözés idejének és mennyiségének ideiglenes változtatására, a szöveges input fájl megváltoztatása nélkül. Emellett a műtrágya kijuttatásának idejét is módosíthatjuk relatív módon, így a felhasználók különböző forgatókönyveket vizsgálhatnak meg, hogy a legmegfelelőbb döntéseket hozhassák meg. A rendszer így lehetővé teszi a mezőgazdasági gyakorlatban felmerülő kihívások és kérdések átfogó és részletes vizsgálatát.

4.3.2. AgroMo Plot (aggregálás és vizualizáció)

A vizualizációs igényeket és a felhasználói elvárásokat az AgroMo Plot (33. ábra), mint grafikon készítő modul szolgálja ki. Ez a modul lehetővé teszi az adatok gyors áttekintését, valamint a szükséges grafikonok és diagramok elkészítését, amelyek elősegítik a döntéshozatali folyamatot és az eredmények értelmezését.

Az AgroMo Plot nem csupán az adatok ábrázolását teszi lehetővé, hanem interaktív eszközöket is kínál a felhasználóknak, hogy mélyebb betekintést nyerjenek az adatokba és könnyedén megtalálják a keresett információkat vagy tendenciákat. Az ilyen típusú integrált vizualizációs modulok hozzájárulnak a rendszer összetett működésének könnyebb megértéséhez és a felhasználói élmény javításához.

	SIMULATION RESULTS:		OBSERVATIONS:					
	Martonvasar_maize	*	data file:	MV_highN.obs				-
	cropland610		-11					
	cropland620		allas:	observation				
	cropland630							
blot	cropland640		CREATE PLOT WITH:		filter	to: all		•
	cropland650		VARI	ABLE	T-STEP	FUNC	PLOT TYPE	7
			Grain Dry Matter (Content	day	identity	line	-
SITE GRID			All Harvested Dry	Matter	day	identity	line	
BASE MAP			Flowering day		day	identity	line	
			Leaf Area Index		day	identity	line	
			Rooting Depth		day	identity	line	
			Cumulative water	stress	day	identity	line	
			Total Soil Carbon	Content	day	identity	line	
•			Water stored in so	bil	day	identity	line	
			Pond water		day	identity	line	
			Water stored in sr	nowpack	day	identity	line	
0			Water stored on c	anopy	day	identity	line	
			Daily Precipitation	n reaching soil	day	identity	line	
			Water going into S	Snowpack	day	identity	line	
			Daily Runoff		day	identity	line	
0			Daily Melt from sr	nowpack	day	identity	line	
			Daily Evaporation		day	identity	line	
			Daily Transpiratio	n	day	identity	line	
		-	Daily Evapotransp	piration	day	identity	line	-
Đ	DELETE SELECTED			CREATE PL	от		EXPOR	T

33. ábra. AgroMo plot.

Az AgroMo Plot modul kifejezetten azokra a felhasználói igényekre lett kialakítva, amelyek a korábbi szimulációk vizuális reprezentációját és összehasonlítását célozzák meg. A modulban lévő funkciók lehetővé teszik a felhasználók számára, hogy egyszerre több szimulációs futtatást válasszanak ki és ábrázoljanak. Több futás párhuzamos ábrázolása (34. ábra) különösen hasznos azokban az esetekben, amikor a kiválasztott szimulációk eltérő beállításokkal, de azonos paraméterek mentén lettek létrehozva. Így a felhasználó azonnal láthatja, hogyan befolyásolhatja akár egyetlen döntés vagy változó a modellezés eredményeit.

Emellett a modulban lehetőség van referenciapontok használatára is. Ezek a pontok gyakran valós mérések vagy korábbi kutatásokból származnak, és fontosak a modell hitelességének és pontosságának ellenőrzésében. Amennyiben egy szimuláció eredménye beleesik a referencia adatpontok által jelölt mérési bizonytalansági tartományba, az egy erős indikátora annak, hogy a modell megfelelően közelíti a valóságot. Az AgroMo Plot ezen kívül lehetőséget biztosít az ismétlések átlagolt eredményeinek ábrázolására és a bizonytalansági tartományok vizualizációjára (ld. 34. ábra), növelve ezzel az adatok értelmezhetőségét és a modell validálásának mélységét.



34. ábra. AgroMo Plot példa.

Az AgroMo Plot modul azon túl, hogy több változót képes egyidejűleg ábrázolni, kiterjedt interaktivitást is biztosít a *Plotly* könyvtár segítségével. A felhasználók nem csak a grafikonok teljes skáláját tekinthetik meg, hanem specifikus időszakokra is ráközelíthetnek, ami lehetővé teszi az adott intervallumok mélyebb elemzését. Továbbá, ha a felhasználók úgy döntenek, hogy bizonyos időszakok nem relevánsak vagy zavaróak a jelenlegi elemzés során, ezeket az időszakokat ki is vághatják az ábrázolásból.

Ezen funkciók különösen hasznosak, amikor a felhasználóknak különböző időintervallumok közötti dinamikákat vagy tendenciákat kell vizsgálniuk. Az adott időszakok kivágása és a grafikonokon való mozgatás képessége egyúttal azt is lehetővé teszi, hogy a felhasználók személyre szabják az ábrázolást, és csak a számukra legfontosabb információkat jelenítsék meg. Az interaktív ábrázolás és ezek a finomító eszközök nem csak a felhasználói élményt fokozzák, hanem segítenek a pontosabb és relevánsabb értelmezésben is.

4.3.3. AgroMo Grid (térben explicit futtatások)

Az AgroMo GRID modul (35. ábra) a komplex mezőgazdasági modellezés modern megközelítését kínálja, amely lehetővé teszi a modellezési adatok térbeli kiterjesztését egy előre definiált rácshálón. Az ilyen típusú rácshálózati modellezés fő előnye, hogy részletes képet ad a területi variabilitásról, így a döntéshozók számára pontosabb információval szolgál a mezőgazdasági tevékenységek optimalizálásához.



35. ábra. Grid futtatások és ensemble szimulációk kezelőfelülete az AgroMo keretrendszerben.

A GRID modul működéséhez szükséges információkat, mint például a különböző pixelek jellemzőit bizonyos időszakokban, a storyline (forgatókönyv) fájlok tárolják. Ezek a fájlok meghatározzák azokat az INI-fájlokat, amelyek az egyes rácspontokra vonatkozó modellfuttatásokat futtatását vezérlik. A felhasználók számára további rugalmasságot biztosít az a képesség, hogy a futtatás előtt módosíthassák a meteorológiai és talajadatokat az egész rácshálóra nézve. Továbbá, módosíthatják azokat az algoritmusokat is, amelyek kulcsszerepet játszanak olyan folyamatok modellezésénél, mint a fotoszintézis vagy az evapotranszspiráció.

Az ensemble jellegű futtatás egy másik innovatív funkció, amely lehetővé teszi a változócsoportok (pl. meteorológia, talaj) összes lehetséges kombinációjának modellezését a teljes rácshálón. Ezzel a módszerrel a felhasználók képesek feltárni a különböző kombinációk hatásait, és jobban megérteni a környezeti tényezők összefüggéseit. Ezen funkcionalitást használtuk az "Input database related uncertainty of BiomeBGCMuSo agro-environmental model outputs" (Fodor et al., 2021) cikkünkben több talajadatbázis és klímaadatbázis kombinációján történő futtatáshoz.

Végül, de nem utolsósorban, a GRID modul nagy mennyiségű adat kezelésére is képes. Az eredmények egy strukturált adatbázisban kerülnek tárolásra, amely megfelelő indexekkel van ellátva a hatékony lekérdezés érdekében. Az adatbázis az alapvető hibakezelést is figyelembe veszi, így a felhasználók megbízható és konzisztens adatokhoz férhetnek hozzá. Ezért minden egyes futtatáshoz egy hiba táblázat is tartozik, ahol a modell hibakódjait tároljuk az összes pixelre vonatkozóan.

A mezőgazdasági modellezés során fontos az adatkezelés hatékonysága és a szimulációs folyamatok gyorsasága. Az AgroMo GRID modul a 10 km-es felbontásban jelenleg 1104 rácsponton való futtatás esetén hatalmas adatmennyiséget kezel. A modell kimeneti adatai tekintélyes terjedelmet öltenek, különösen, ha a komplexitást figyelembe vesszük: 10 kimeneti változó, 100 év, naponta frissített adatok. Egyetlen szcenárió futtatása során 1104*100*365*10, azaz több mint 4 milliárd adatpont keletkezik, ami körülbelül 768 MB tárhelyet igényel.

Az ensemble modellezés azonban további komplexitást ad a rendszerhez. Tíz különböző RCM/GCM modellpárral, páronként két szcenárióval, valamint három különböző talajadatbázissal, a

teljes ensemble modell futtatási száma meghaladja a 66 000 futtatást. Ezen futtatások összességében körülbelül 46 GB tárhelyigényt eredményeznek.

Tekintettel a modell futási idejére, egyetlen futtatás 1 másodpercig tart, és a teljes ensemble modell futtatás közel 18.4 órát vesz igénybe egyetlen számítási szálon. Ez jelentős idő, amely alatt a rendszernek képesnek kell lennie az adatok hatékony kezelésére és tárolására anélkül, hogy a teljesítmény romlana.

Ezért az AgroMo a legmodernebb konkurens-programozási stratégiákat szerint több szálon aszinkron módon futtatja a különböző rácsokon a modelleket. Szerencsére, mivel a rácsok egymástól függetlenek a párhuzamosítás hatásfoka közel van az elméleti maximumhoz. Olyannyira, hogy a keletkezett adatok adatbázisba írása a fenn említett mennyiségű adat esetén lassabb, mint a modellek futtatása maga. Ennek következtében a fenn említett 18,4 óra helyett 1 órán belül végzünk a futtatással egy 24 szálat alkalmazni tudó számítógép esetén.

4.3.4. AgroMo Query (lekérdezések egyszerűsítése, riportkészítés)

Az AgroMo Grid segítségével elkészült táblák kiértékelésére használhatunk előre elkészített sablonokat, amelyeket a metanyelv segítségével készítettünk (lásd. 3.7.4. fejezet). Az AgroMo Query modulban (36. ábra), amely egy lapon szerepel a Griddel, ezek közül egy listából választhatunk lekérdezést, amelynek adhatunk nevet és egy rövid leírást is (ezek később a térkép készítő modulban lesznek felhasználva).

	τέρκέρ μυὄζ	SCC7FC 11# #		1 (2. Star -) - (-) [2 ******* [/ ******	41- +1 1- 2 1 (-) [T T]	:
	{5: FORESEE40)} adatforrás al	eg (10= <mark>(1: 0)</mark> apján	(2: attaga) a() (3: "months") - (4: "mon	ths"} попарокга a(z) [1-1] i	idoszakban a(z)
	TÉRKÉP KV (CSOPORT)HATÁ	SA A NPP-ra	CELLÁNKÉNT a	(z) [T-T] időszakra a(z) <mark>(1</mark> :	*tables*} szimuláció alapjá	án
	TÉRKÉP (CSC alapján	OPORT) ÁTLAG F	PROFIT a(z) [T <mark>-T] i</mark> dőszakba	n a(z) <mark>{1: *tables*}</mark> szimulá	ció és a(z) <mark>{2: 0.}</mark> ökonómia	ai forgatókönyv
0	TÉRKÉP TER	MÉS- <mark>B</mark> IZONYTAI	L <mark>AN</mark> SÁG CELL	LÁNKÉNT a(z) [T-T] időszakban a(z) <mark>{1: *ta</mark>	<mark>bles*}</mark> szimuláció alapján	
	LEKÉRDEZÉS AL	IAS:		LEÍRÁS:			
L L		[tól-ig]:	•	- {5}:	NA	•	LEKÉRDE
(1):	NA	[tól-ig]:	•	₹ {5}:₹ {6}:	NA	• •	LEKÉRDE
(1):	NA NA	[tól-ig]:	• •	 {5}: {6}: {7}: 	NA NA NA	ا به د این ا این ا	LEKÉRDE
(1): (2): (3):	NA NA	[tól-ig]:	•	 {5}: {6}: {7}: {8}: 	NA NA NA NA	• • • • • • • • • • • • • • • • • • •	LEKÉRDE JELENTÉS g mutat & ment

36. ábra. Alekérdezések kezelőfelülete az AgroMo keretrendszerben.

Kapcsos zárójelben látjuk a helyettesíthető értékeket az aktuális választással, amelyet az opciók kiválasztásával lehet megváltoztatni. Két típusú lekérdezés létezik: az, amelyet térképre szánunk

(map) valamint az általános típusú (table). Ha a "LEKÉRDEZ" gombra kattintunk (query), létrejön a megfelelő SQL lekérdezés, valamint az adat is, amely a lekérdezés végeredménye. A "JELENTÉS" gombra kattintva a lekérdezés eredménye egy MS Excel táblában is eltárolódik. A "TÉRKÉP" gombon keresztül megjelenik a térképen ábrázoló modul.

Egy konkrét példával demonstrálom az eszköz hatékonyságát. Az eszköz használatával a következő lekérdezések eredményei lettek közölve az "Input database related uncertainty of BiomeBGCMuSo agro-environmental model outputs" (Fodor et al., 2021) című szakcikkben, amelynek megszületéséhez munkám jelentős mértékben hozzájárult:

- A globálsugárzás átlagos napi intenzitásának, az átlagos napi középhőmérsékletnek és a csapadékösszegnek a vegetációs időszakra (április-október) vonatkozó különbségei 1981-2010 között a vizsgált meteorológiai adatbázisok alapján.
- A talajprofil felső 100 cm-es rétegének növény által felvehető vízkészletének különbségeit a vizsgált adatbázisok alapján számítottuk ki: EU-PTF, HU-PTF és US-PTF³⁴. Az EU-SHG adatbázist használtuk viszonyítási alapként az összehasonlításokhoz.
- Rácscellák, ahol a növénytermesztési, gyep- és erdőterület aránya meghaladja a 20%ot.

A fenti példákból is látható, hogy pár kattintásból AgroMo nagy komplexitású lekérdezések elvégzésére képes a rendszer (a fenti lekérdezésekből az AgroMo Map ábrát is készített).

4.3.5. AgroMo Map (térképek készítése)

Az AgroMo Query modulban történt lekérdezések egyik lehetséges kimenete térképen ábrázolható. Az AgroMo Map (37. ábra) erre biztosít felhasználóbarát lehetőséget. A felületen kiválaszthatjuk az ábrázolni kívánt adatforrást, amely egy előző lekérdezés generált SQL fájlja, vagy saját magunk által kézzel alkotott SQL fájl. A lekérdezést követően egy azonos nevű *csv* kiterjesztésű fájl keletkezik, ami közvetlenül ábrázolható. Amennyiben ez a fájl már létezik (az SQL fájl az AgroMo query eredménye), újabb lekérdezés már nem történik, a megfelelő adatokat közvetlenül ábrázolhatjuk. Ezzel az SQL-hez értő felhasználó akár ki is hagyhatja az előző lekérdező modult és saját maga is megírhatja közvetlenül a lekérdezést, így biztosítva egyszerre a szükséges rugalmasságot és kényelmet a felhasználóknak.

³⁴ Talajadatbázisok rövidítései; részletek lásd a Fodor et al. (2021) szakcikkben.



37. ábra. AgroMo Map létrehozása a grafikus felhasználói felületen (példa).



38. ábra. A modellezett NPP aggregált éves bizonytalanságának térbeli eloszlása az éghajlati (fent) és a talaj (lent) adatbázisok megválasztásának függvényében. 2007 az egyik leginkább aszályos év volt, 2010 pedig a legcsapadékosabb év a vizsgált időszakban (1981-2010) (Fodor et al. (2021) 6. ábrája).

A térképezéshez beállítható, hogy szeretnénk-e szélességi/hosszúsági-vonalakat, látszódjon-e az országhatár, valamint a paletta (összesen 11-féle), és a színskála felbontása, továbbá az ábrának adhatunk címet is. A térképeket automatikusan 600 DPI felbontásban png formátumban tárolja a rendszer. A felület alján mindig a két utolsó ábra látható, így összehasonlításra is hatékonyan lehet használni a rendszert.

Az AgroMo Map használhatóságát kiválóan illusztrálja, hogy ezzel az eszközzel készült a Fodor et al. (2021) című szakcikk összes térképes összehasonlító ábrája, amelyből az egyik a 38. ábrán látható.

4.3.6. AgroMo paraméter optimalizáció (ParAna)

A 39. ábrán látható ParAna felületen keresztül elérhető az RBBGCMuso kalibrációs modulja. Itt ki tudunk választani egy munkakönyvtárat, amelyben a kalibrációt leíró fájlok találhatók (lásd. 3.7.6 fejezet). A megfigyelési fájlok szabadon választhatóak a felületről. A felületen történik a Monte Carlo iterációk számának megadása, valamint a CIRM iterációs lépések számának, vagy a sikerességi arány megadásának lehetősége (ha az elvárt sikerességi arányt korábban elérjük, nem hajtjuk végre az összes CIRM iterációt). Továbbá szabadon változtathatóak a *constraint.json*, valamint a kalibrálandó paramétereket leíró *set* fájl is.

A felületen a kalibrációt elvégezve a mért értékeket a szimulált értékekkel összevető grafikon jelenik meg, amely segítségével vizuálisan is megvizsgálhatjuk, hogy a kalibrálás előtti paraméterbeállításokhoz képest a kalibrálás mekkora pontosság-növekedéssel járt. A változásokat a grafikon mellett egy táblázatban szereplő RMSE/MAE/R² értékek is szemléltetik. Az összehasonlítás alapjául szolgáló értékek, akárcsak a metrikák egy results.xlsx nevű Excel fájlba kerülnek.

A jelenlegi megvalósítás egyelőre nem tartalmazza az RBBGCMuso lokális és globális érzékenység-elemzési eszközeit. A megvalósítás szándékát jelzi, hogy a grafikus elemek már elkészültek (ld. "Parameter Sweep" és "Sensitivity Analysis" a 39. ábrán).



39. ábra. AgroMo Paraméter Analízis grafikus felülete.

4.4. Projekciók kukoricára

A fent bemutatott eszközök (RBBGCMuso, CIRM, AgroMo, storyline, SQLite adatbázis, SQL lekérdező metanyelv) együttes használata lehetővé teszi, hogy a rendelkezésünkre álló legkorszerűbb klímaprojekciókat társítsuk a legújabb Biome-BGCMuSo modellverzióval, és a hazai mezőgazdasági termelés kapcsán hatásvizsgálatokat készítsünk. Kiemelendő, hogy ennek elkészítése általában több hónapos előkészítő munkát igényel, tekintve az adatbázisok diverz jellegét, hozzáférhetőségi korlátait, formátumát, a hozzájuk kapcsolódó problémákat (pl. a klímaadatok bias-korrekciója kapcsán), és a vizsgálat térben explicit jellegét.

Az AgroMo szoftver grafikus felületén egy rövid és megismételhető munkafolyamattal ensemble típusú hatásvizsgálatot végeztem a hazai kukoricatermelés jövőben várható változásának számszerűsítésére.



40. ábra. A mezőgazdasági hatásvizsgálat eredménye az 5-5 kiválasztott, FORESEE alapú klímamodell ensemble eredménye alapján teljes Magyarországra. A fekete pontok a KSH országos szintű megfigyeléseit reprezentálják (csak azt az időszakot ábrázoltam, amikor a műtrágyázás mennyisége hasonló volt az általunk feltételezett mennyiséggel). A világosszürke sávok a bizonytalanságot reprezentálják. A piros és zöld görbék az ún. LOESS módszerrel simított adatokat reprezentálják a könnyebb értelmezés érdekében.

A 40. ábrán idősor formájában láthatjuk a szimuláció eredményét, országos szintre aggregálva. Az ensemble modellezés szabályai szerint ez a jelenlegi ismereteink szerinti legvalószínűbb projekció mind a RCP4.5, mind az RCP8.5 esetén, ahol RCP4.5 esetén a 2041-2060 közötti időszakra átlagosan 56 mm-rel kevesebb éves csapadékösszegre, valamint 1,17 °C-kal magasabb hőmérsékletre lehet számítani, RCP85 esetén pedig 38,7 mm-el kevesebb átlagos éves csapadékösszegre Fontos kiemelni, hogy az eredmények csak a felhasznált feltételezések keretei közt értelmezhetőek (vagyis változatlan kezelés mellett).

Az ábra egy érdekes mintázatot mutat, és nem intuitív trajektóriákat sugall az időjárás termelésre gyakorolt hatása kapcsán. A közeli jövőben egy fokozatos csökkenést lehet valószínűsíteni, ami az RCP4.5 és 8.5 kapcsán nem tér el egymástól lényegesen. 2040 és 2070 között az RCP8.5 csekély mértékben, de magasabb termésmennyiséget mutat, mint az RCP4.5. Az RCP4.5 szerint a 2050 utáni időszakban a tendencia megváltozik, és a kukorica termésmennyiség csökkenés lassú növekedésbe megy át, és 2100-ig majdnem eléri a kezdeti szintet. Ezzel szemben az RCP8.5 pesszimista szcenárió szerint egy kis megtorpanástól eltekintve csökkenésre lehet számítani, ami a század végén igen drasztikus terméskiesést jelez. Az ensemble eredmények szerint a 2001-2020 referencia-időszakhoz viszonyítva a 2041-2060-as időszakban átlagosan -15% (-1,17 t/ha; RCP4.5)

90

illetve -13% (-1.01 t/ha; RCP8.5), míg a 2081-2100 időszakra nézve -10% (-0.77; RCP4.5) illetve - 28% (-2.20; RCP8.5) termésátlag változás várható.

Fontosnak tartom megjegyezni, hogy számos tanulmány csak egyetlen, vagy csak igen kis számú klímaprojekció alapján közöl becsléseket a jövőre nézve. Ez sok esetben elégtelen, mert ha a sokaságból egy extrémebb modellt választunk, az eredmény megtévesztő lehet.



41. ábra. Modellezett százalékos termésmennyiség változás a 2001-2020 közötti időszak átlagához viszonyítva a 2041-2060, illetve 2081-2100 időszakra nézve. A felső sorban az RCP4.5 forgatókönyvek átlagát, alsó sorban az RCP8.5 forgatókönyvek átlagát használtam (a szürke pixeleken nem termesztenek kukoricát).

Térbeli vizsgálat céljából a 2001-2020 közötti referencia időszakhoz képest térképesen ábrázoltam (41. ábra) a változás mértékét, az 5-5 db klímaprojekció átlagaként. Ahogy az idősoros ábra alapján már láttuk, a termésmennyiség csökkenése várható a jövőben. A változás mértéke az országban nem egyenletes, és jól kivehető térbeli mintázatokat fedezhetünk fel. A 2041-2060 időszakban az RCP4.5 szerint átlagosan 15,3%-os csökkenésre lehet számítani (az 5., illetve 95. percentilis a pixel szintű adatok alapján 24%, illetve 1.9% csökkenés). Az RCP8.5 alapján az átlagos változás ugyan csekély mértékben, de kisebb (átlag: -13,6%, 5., illetve 95. percentilis rendre -22,6 és +1,6%). A távoli jövőben a becslések közötti olló szétnyílik. A 2081-2100 időszakban az RCP4.5, vagyis az optimistább szcenárió szerint az átlagos csökkenés csak 10,8% (az 5., illetve 95. percentilis 18,7% csökkenést, illetve +4.4% növekedést mutat). Az RCP8.5 szerint az átlagos termésátlag csökkenés 28,6% (az 5., illetve 95. percentilis -34,5%, illetve -17,5%).

-	RCP	4.5	RCP8.5		
	2040-2060	2081-2100	2040-2060	2081-2100	
Baranya	-20	-12	-18	-28	
Borsod-Abaúj-Zemplén	-21	-16	-18	-33	
Bács-Kiskun	-11	-6	-9	-26	
Békés	-12	-9	-11	-25	
Csongrád	-17	-11	-17	-30	
Fejér	-22	-15	-21	-33	
Győr-Moson-Sopron	-6	-3	-2	-22	
Hajdú-Bihar	-13	-11	-12	-29	
Heves	-14	-13	-14	-28	
Jász-Nagykun-Szolnok	-15	-12	-16	-28	
Komárom-Esztergom	-12	-6	-9	-27	
Nógrád	-8	-5	-8	-23	
Pest	-19	-15	-18	-29	
Somogy	-18	-12	-13	-32	
Szabolcs-Szatmár-Bereg	-6	-1	-2	-22	
Tolna	-23	-17	-21	-32	
Vas	-5	1	-2	-19	
Veszprém	-11	-3	-7	-25	
Zala	-12	-6	-11	-27	

8. táblázat. A kukorica termésmennyiségének projektált százalékos változásai megyei szinten, az 5-5 klímaprojekció alapján a jelenlegi termésszinthez képest, RCP4.5 és RCP8.5 szcenáriók alapján a közeli jövőre, illetve a távoli jövőre.

Ahhoz, hogy részletesebb információt kapjunk, megyei szinten is aggregáltam a várható változás mértékét a két RCP szcenáció alapján (8. táblázat). A megyei szintű adatok alapján az RCP4.5 szerint a közeli jövőben Tolna, Baranya, Fejér és Borsod-Abaúj Zemplén megyék esetén lesz a legjelentősebb hatás, ami érinti a jelenleg legmagasabb termésátlaggal rendelkező területeket (Tolna, Baranya, Fejér megyék). A távoli jövőben is hasonlóan alakul a leginkább érintett megyék listája. Az RCP8.5 szerint a közeli jövőben Tolna és Fejér megyékben lehet 20%-nál nagyobb csökkenés, de Pest, Csongrád, Baranya és Borsod-Abaúj Zemplén megyék is komoly, időjárás által keltett terméskieséssel számolhatnak. A távoli jövőben a várható hatások még jelentősebbek; ekkor csupán Vas megyében jelenik megy egy 20% alatti csökkenés. Borsod-Abaúj-Zemplén, Fejér, Somogy és Tolna megye esetén a terméskiesés 30%-nál is több. További vizsgálatokra és alternatív szcenáriók konstruálására van szükség ahhoz, hogy a lehetséges károk csökkentésének mértékét meg lehessen becsülni. A 2022-es év nyarán előfordult extrém aszály rámutat arra, hogy stratégiai jelentőségű a megfelelő szintű felkészülés az éghajlatváltozás kockázatai kapcsán.

A projekciók eredményeit összevetettem a korábbi, Magyarország teljes területére vonatkozó hasonló hatásvizsgálatok eredményeivel. Kizárólag folyamatorientált modellel történő termésprojekciókal foglalkoztam. Az összes itt felsorolt tanulmány feltételezi, hogy a növénytermesztési szokások nem fognak változni, tehát az eredmények összevethetők a fent közöltekkel. Fodor & Pásztor (2010) a 4M modell segítségével a 2100-ra vonatkozó időjárási szcenárióval élve (3. táblázat a vonatkozó tanulmányban) átlagosan 7% terméshozam-csökkenést vetítettek előre (a 2000-es évekhez képest). A Fodor et al. (2014) tanulmány ugyanerre az időszakra már GCM-alapú klímaprojekciót használt, amelyet egy regionális klímamodellel dinamikusan leskálázott. A klímaprojekció az A1B (Dobor et al., 2015) klímaforgatókönyv alapján készült, amit az RCP6.0szcenáriónak lehet megfeleltetni (van Vuuren & Carter, 2014). A tanulmány szintén a 4M modellel szimulálta a kukorica terméshozamot. Eredményei az előző tanulmánynál több mint háromszor nagyobb (-24%) terméshozam változást mutatnak a századfordulóra (ami az én tanulmányom RCP8.5-ös 2081-2100 időszakra vonatkozó 28% csökkenéséhez áll legközelebb.). Jelentős előrelépés volt a témában Dobor (2016) disszertációja, amelyben a szerző két növénytermesztési modellt is alkalmazott (a 4M-met, valamint a Biome-BGCMuSo korábbi, 4.1.1-es verzióját), 10 GCM/RCM párral az A1B forgatókönyvre vonatkozóan. A klímamodellekre vonatkozó multi-modell átlag terméscsökkenés Biome-BGCMuSo esetén 2050-ig 9%, 2100-ig pedig 11% volt, a 4M esetén pedig rendre 6%, illetve 24%. Fontos megjegyeznünk itt, hogy a referencia (1986-2013), a közeli (2021-2050) és távoli jövő (2071-2100) időszakai jelentősen eltérnek jelen dolgozatban szereplő időszakoktól, mind hosszban, mint kezdőévet tekintve. Webber et al. (2018) szintén 30 éves időszakra vonatkoztatva közölt Magyarországra, kukoricára vonatkozó terméshozam változást, azonban itt 1981-2010 volt a referencia időszak és 2040-2069 a vizsgált időszak. Erre az időszakra Webber et al. (2018) 5 GCM-et használt RCP2.6, RCP4.5, illetve RCP8.5 forgatókönyvekkel, valamint 6 növénytermesztési modellt (amelyek közül az egyik modell a 4M volt). A növénytermesztési, illetve meteorológiai modellek átlagaiból a várható termésmennyiség változás rendre -9% (RCP2.6), -16% (RCP4.5), illetve -22% (RCP8.5).

Összességében valamennyi Magyarországra vonatkozó eddigi tanulmány terméscsökkenést projektál elő mind a század közepére, mind a század végére, aminek oka, hogy a kukorica rendkívül érzékeny a hőstresszre és szárazságra (Olesen et al., 2011). A tanulmányok többsége egy forgatókönyvet használt (A1B), míg Webber et al. (2018), valamint ezen tanulmány több RCP forgatókönyvet vettek figyelembe. Dolgozatomban a Webber et al. (2018) tanulmánnyal ellentétben viszont a század közepén az RCP8.5 szcenárió kisebb terméscsökkenéssel járhat, mint az RCP4.5. A századfordulóra azonban a dolgozatomban megjelenő 24%-os terméscsökkenés az összes eddigi projekció terméscsökkenésénél nagyobb mértékű.

Az eredmények rámutatnak a korszerű klímaprojekció sokaság alkalmazásának fontosságára, és a modell nem intuitív válaszára a különböző irányú hatásokra.

5. Összefoglalás

Doktori munkám során a hazai fejlesztésű Biome-BGCMuSo folyamatorientált biogeokémiai modell alkalmazását támogató keretrendszereket fejlesztettem ki és alkalmaztam döntéshozói kontextusban. A fejlesztések vezérlőelve az volt, hogy a modell-adat fúzió minél több szegmensére nyújtsak konkrét, a gyakorlatban is alkalmazható megoldást. Mivel minden általam készített szoftver nyílt forráskódú, ráadásul GitHub-on szabadon elérhető, ezért az érdeklődő kutatók minden itt nem közölt részlethez hozzáférnek.

A munkám részeként egy új eljárást dolgoztam ki a folyamatorientált modellek optimalizálására (vagyis a modellparaméterek bizonytalanságának minimalizálására), amely túlmutat a hagyományos valószínűségi módszereken. Az általam javasolt CIRM módszer leginkább a hagyományos valószínűségi módszerek és egy jól értelmezhető gépi tanulási módszer kombinációjaként írható le. Legjobb tudásom szerint a szakirodalomban nem található hasonló megközelítés. A munkám során pontszerű állomási adatok felhasználásával optimalizáltam a Biome-BGCMuSo-t a CIRM segítségével, majd a modellt megyei (NUTS3) szinten validáltam.

A bemutatott RBBGCMuso szoftvercsomag elsősorban a kutatói közösséget célozza meg, mivel valamennyi programozói ismeretet igényel. A szoftvert ma már nem csak magyar kutatók használják. Nemzetközi együttműködések keretében horvát, szlovák, cseh, új-zélandi, brazil és kínai kutatók is alkalmazzák. A Biome-BGCMuSo modellcsalád funkciógazdagságát tekintve nem létezik hasonló fejlesztés a tudományos közösségben.

A saját fejlesztésű AgroMo rendszer összegzi a munkámat. Ebben letisztult formában, grafikus felhasználói felület mögé rejtve minden bemutatott komponens szerepel: a Biome-BGCMuSo, az RBBGCMuso, a FORESEE adatbázis, a DoSoReMI talajadatbázis, az SQL alapú természetes query-nyelv, a storyline funkcionalitás és a párhuzamosítás. Munkám eredményét bárki letöltheti a nyílt elérésű GitHub felületről (<u>https://github.com/hollorol/AgroMo</u>).

A dolgozat végén bemutatott esettanulmány a jelenleg rendelkezésre álló legkorszerűbb alapadatokat használja, és reprodukálható módon vetíti előre a kukorica terméshozamát az RCP4.5, illetve RCP8.5 forgatókönyvek alapján. Mivel az agrotechnika kapcsán azzal a feltételezéssel éltünk, hogy nem lesz változás, ezért az eredmények a változó időjárás és a növekvő légköri CO₂ koncentráció együttesen kifejtett hatását jelenítik meg. A projekciók alapján, amennyiben az emberi beavatkozás jellege nem változik, úgy a terméshozam csökkenése várható. Az eredmények rámutatnak további forgatókönyvek konstruálásának fontosságára, ahol klímaadaptív mezőgazdasági művelési módok bevezetését is számításba kell venni, mint pl. az öntözés, a vetési idő módosulása, és szárazságtűrő hibridek bevezetése.

6. Köszönetnyilvánítás

A doktori kutatás részben a Széchenyi 2020 program keretében, Magyarország Kormánya és az Európai Regionális Fejlesztési Alap támogatásával (GINOP-2.3.2-15-2016-00028 számon) valósult meg. A kutatást az Éghajlatváltozás Nemzeti Multidiszciplináris Laboratórium (RRF-2.3.1-21-2022-00014 számú) projekt is támogatta.

Munkámat továbbá az Innovációs és Technológiai Minisztérium a Nemzeti Kutatási Fejlesztési és Innovációs Alapból a TKP2021-NKTA pályázati program keretén belül támogatta (TKP2021-NKTA-06 számon).

Köszönöm a GINOP projekt résztvevő kutatóinak, hogy támogatták az AgroMo szoftver bizonyos komponenseinek a kifejlesztését (Kristóf Erzsébet: vizualizáció; Kis Anna: SQL; Hidy Dóra: Biome-BGCMuSo alapmodell).

Köszönettel tartozom témavezetőimnek: Barcza Zoltánnak, Fodor Nándornak is, akik nem csak a tudományos életutamat egyengették, hanem segítettek jobb emberré is válnom. Nagy türelmük és folyamatos támogatásuk nélkül a munkám még csak meg sem kezdődhetett volna.

Köszönöm barátaimnak és családomnak, elsősorban feleségemnek, Gergő Viviennek a folyamatos támogatást és türelmet.

7. Irodalomjegyzék

- Ainsworth, E. A., & Long, S. P. (2021). 30 years of free-air carbon dioxide enrichment (FACE): What have we learned about future crop productivity and its potential for adaptation? *Global Change Biology*, *27*(1), 27–49. DOI: 10.1111/gcb.15375
- Albergel, C., Dutra, E., Munier, S., Calvet, J.-C., Munoz-Sabater, J., de Rosnay, P., & Balsamo, G. (2018). ERA-5 and ERA-Interim driven ISBA land surface model simulations: Which one performs better? *Hydrology and Earth System Sciences*, *22*(6), 3515–3532. DOI: 10.5194/hess-22-3515-2018
- Anderson, R., Keshwani, D., Guru, A., Yang, H., Irmak, S., & Subbiah, J. (2018). An integrated modeling framework for crop and biofuel systems using the DSSAT and GREET models. *Environmental Modelling & Software*, *108*, 40–50. DOI: 10.1016/j.envsoft.2018.07.004
- Angulo, C., Rötter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in Europe.
 Agricultural and Forest Meteorology, *170*, 32–46. DOI: 10.1016/j.agrformet.2012.11.017
- Bagnara, M., Silveyra Gonzalez, R., Reifenberg, S., Steinkamp, J., Hickler, T., Werner, C., Dormann,
 C. F., & Hartig, F. (2019). An R package facilitating sensitivity analysis, calibration and forward simulations with the LPJ-GUESS dynamic vegetation model. *Environmental Modelling & Software*, *111*, 55–60. DOI: 10.1016/j.envsoft.2018.09.004
- Basso, B., Liu, L., & Ritchie, J. T. (2016). A comprehensive review of the CERES-wheat, -maize and-rice models' performances. *Advances in Agronomy*, *136*, 27–132. DOI: 10.1016/bs.agron.2015.11.004
- Bellman, R., 1957. Dynamic Programming. Princeton University Press. ISBN: 0-691-07951-X
- Bengtsson, H. (2021). A Unifying Framework for Parallel and Distributed Processing in R using Futures. *The R Journal*, *13*(2), 208–227. DOI: 10.32614/RJ-2021-048

- Beven, K., & Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes*, *28*(24), 5897–5918. DOI: 10.1002/hyp.10082
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1), 11–29. DOI: 10.1016/S0022-1694(01)00421-8
- Bilionis, I., Drewniak, B. A., & Constantinescu, E. M. (2015). Crop physiology calibration in the CLM. *Geoscientific Model Development*, *8*(4), 1071–1083. DOI: 10.5194/gmd-8-1071-2015
- Bloom, A. A., & Williams, M. (2015). Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological" common sense" in a model–data fusion framework. *Biogeosciences*, *12*(5), 1299–1315. DOI: 10.5194/bg-12-1299-2015
- Bonan, G. B., Levis, S., Kergoat, L., & Oleson, K. W. (2002). Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, *16*(2), 5–1. DOI: 10.1029/2000GB001360
- Bond-Lamberty, B., Gower, S.T., Ahl, D.E., & Thornton, P.E., (2005). Reimplementation of the Biome-BGC model to simulate successional change. *Tree Physiology*, 25, 413–424. DOI: 10.1093/treephys/25.4.413
- Braswell, B. H., Sacks, W. J., Linder, E., & Schimel, D. S. (2005). Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology*, *11*(2), 335–355. DOI: 10.1111/j.1365-2486.2005.00897.x
- Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi,
 P., & Burger, P. (2003). An overview of the crop model STICS. *European Journal of Agronomy*, *18*(3–4), 309–332. DOI: 10.1016/S1161-0301(02)00110-7
- Canadell, J. G., Monteiro, P. M. S., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V.,
 Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P. K., Piao, S., Rogelj, J.,
 Syampungani, S., Zaehle, S., & Zickfeld, K. (2021). Global Carbon and other Biogeochemical
 Cycles and Feedbacks. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S.

Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 673–816). Cambridge University Press. DOI: 10.1017/9781009157896.007

- Sierra C. A., Markus M., & Susan E. Trumbore. (2012). Models of soil organic matter decomposition: The SoilR package, version 1.0. *Geoscientific Model Development*, 5(4), 1045–1060. DOI: 10.5194/gmd-5-1045-2012
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2017). *shiny: Web Application Framework for R.* https://CRAN.R-project.org/package=shiny
- Chapin, F.S., Woodwell, G.M., Randerson, J.T., Rastetter, E.B., Lovett, G.M., Baldocchi, D.D., Clark,
 D.A., Harmon, M.E., Schimel, D.S., Valentini, R., Wirth, C., Aber, J.D., Cole, J.J., Goulden,
 M.L., Harden, J.W., Heimann, M., Howarth, R.W., Matson, P.A., McGuire, A.D., Melillo,
 J.M., Mooney, H.A., Neff, J.C., Houghton, R.A., Pace, M.L., Ryan, M.G., Running, S.W.,
 Sala, O.E., Schlesinger, W.H., Schulze, E.-D., (2006). Reconciling Carbon-cycle Concepts,
 Terminology, and Methods. *Ecosystems*, 9, 1041–1050. DOI: 10.1007/s10021-005-0105-7
- Churkina, G., Tenhunen, J., Thornton, P., Falge, E. M., Elbers, J. A., Erhard, M., Grünwald, T., Kowalski, A. S., Rannik, Ü., & Sprinz, D. (2003). Analyzing the ecosystem carbon dynamics of four European coniferous forests using a biogeochemistry model. *Ecosystems*, *6*(2), 0168–0184. DOI: 10.1007/s10021-002-0197-2
- Ciais, P., Dolman, A.J., Bombelli, A., Duren, R., Peregon, A., Rayner, P.J., Miller, C., Gobron, N., Kinderman, G., Marland, G., Gruber, N., Chevallier, F., Andres, R.J., Balsamo, G., Bopp, L., Bréon, F.-M., Broquet, G., Dargaville, R., Battin, T.J., Borges, A., Bovensmann, H., Buchwitz, M., Butler, J., Canadell, J.G., Cook, R.B., DeFries, R., Engelen, R., Gurney, K.R., Heinze, C., Heimann, M., Held, A., Henry, M., Law, B., Luyssaert, S., Miller, J., Moriyama, T., Moulin, C., Myneni, R.B., Nussli, C., Obersteiner, M., Ojima, D., Pan, Y., Paris, J.-D., Piao, S.L.,

Poulter, B., Plummer, S., Quegan, S., Raymond, P., Reichstein, M., Rivier, L., Sabine, C., Schimel, D., Tarasova, O., Valentini, R., Wang, R., Van Der Werf, G., Wickland, D., Williams, M., Zehner, C. (2014). Current systematic carbon-cycle observations and the need for implementing a policy-relevant carbon observing system. *Biogeosciences*, *11*(13), 3547–3602. DOI: 10.5194/bg-11-3547-2014

- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., & Carrara, A. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, *437*(7058), 529–533. DOI: 10.1038/nature03972
- Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377–387. DOI: 10.1145/362384.362685
- Cox, P. M. (2019). Emergent Constraints on Climate-Carbon Cycle Feedbacks. *Current Climate Change Reports*, 5(4), 275–281. DOI: 1007/s40641-019-00141-y
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., & Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, *15*(3), 183–203. DOI: 10.1007/s003820050276
- De Oliveira Ferreira Silva, C., Heriberto De Castro Teixeira, A., & Lilla Manzione, R. (2019). agriwater: An R package for spatial modelling of energy balance and actual evapotranspiration using satellite images and agrometeorological data. *Environmental Modelling & Software*, *120*, 104497. DOI: 10.1016/j.envsoft.2019.104497
- De Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., van Kraalingen, D., Supit, I., van der Wijngaart, R., & van Diepen, K. (2019). 25 years of the WOFOST cropping systems model. *Agricultural Systems*, *168*, 154–167.
- Di Castro, F., & Bertini, E. (2019). Surrogate decision tree visualization interpreting and visualizing black-box classification models with surrogate decision tree: 2019 Joint ACM IUI Workshops, ACMIUI-WS 2019. *CEUR Workshop Proceedings, 2327.* http://www.scopus.com/inward/record.url?scp=85063203196&partnerID=8YFLogxK

- Di Paola, A., Valentini, R., & Santini, M. (2016). An overview of available crop growth and yield models for studies and assessments in agriculture. *Journal of the Science of Food and Agriculture*, 96(3), 709–714. DOI: 10.1002/jsfa.7359
- Di Vittorio, A. V., Anderson, R. S., White, J. D., Miller, N. L., & Running, S. W. (2010). Development and optimization of an Agro-BGC ecosystem model for C4 perennial grasses. *Ecological Modelling*, *221*(17), 2038–2053. DOI: 10.1016/j.ecolmodel.2010.05.013
- Dobor, L. (2016). Az éghajlatváltozás lehetséges hatásai a mezőgazdasági szénmérlegre és produktivitásra Magyarországon [Doktori értekezés]. Eötvös Loránd Tudományegyetem, Meteorológiai Tanszék.
- Dobor, L., Barcza, Z., Hlásny, T., Havasi, Á., Horváth, F., Ittzés, P., & Bartholy, J. (2015). Bridging the gap between climate models and impact studies: The FORESEE Database. *Geoscience Data Journal*, *2*(1), 1–11. DOI: 10.1002/gdj3.22
- Ewert, F., Rötter, R.P., Bindi, M., Webber, H., Trnka, M., Kersebaum, K.C., Olesen, J.E., Van Ittersum, M.K., Janssen, S., Rivington, M., Semenov, M.A., Wallach, D., Porter, J.R., Stewart, D., Verhagen, J., Gaiser, T., Palosuo, T., Tao, F., Nendel, C., Roggero, P.P., Bartošová, L., & Asseng, S., (2015). Crop modelling for integrated assessment of risk to food production from climate change. *Environmental Modelling & Software*, 72, 287–303. DOI: 10.1016/j.envsoft.2014.12.003
- Fodor, N. (2006). 4M-Software for Modelling and Analysing Cropping Systems. *Journal Of Universal Computer Science* 12, 1196–1207. DOI: 10.3217/JUCS-012-09-1196
- Fodor, N., & Pásztor, L. (2010). The agro-ecological potential of Hungary and its prospective development due to climate change. *Applied Ecology and Environmental Research*, *8*(3), 177–190.
- Fodor, N., Pásztor, L., & Németh, T. (2014). Coupling the 4M crop model with national geo-databases for assessing the effects of climate change on agro-ecological characteristics of Hungary. *International Journal of Digital Earth*, *7*(5), 391–410. DOI: 10.1080/17538947.2012.689998

- Fodor, N., Pásztor, L., Szabó, B., Laborczi, A., Pokovai, K., Hidy, D., Hollós, R., Kristóf, E., Kis, A.,
 & Dobor, L. (2021). Input database related uncertainty of Biome-BGCMuSo agroenvironmental model outputs. *International Journal of Digital Earth*, *14*(11), 1582–1601.
 DOI: 10.1080/17538947.2021.1953161
- Foley, J. A., Levis, S., Costa, M. H., Cramer, W., & Pollard, D. (2000). Incorporating dynamic vegetation cover within global climate models. *Ecological Applications*, *10*(6), 1620–1632.
 DOI: 10.1890/1051-0761(2000)010[1620:IDVCWG]2.0.CO;2
- Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., & Haxeltine, A. (1996).
 An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles*, *10*(4), 603–628. DOI: 10.1029/96GB02692
- Free Software Foundation (2023). *Gawk* [Computer software]. Free Software Foundation. https://www.gnu.org/software/gawk/
- Friedlingstein, P., Jones, M.W., O'Sullivan, M., Andrew, R.M., Bakker, D.C.E., Hauck, J., Le Quéré,
 C., Peters, G.P., Peters, W., Pongratz, J., Sitch, S., Canadell, J.G., Ciais, P., Jackson, R.B.,
 Alin, S.R., Anthoni, P., Bates, N.R., Becker, M., Bellouin, N., Bopp, L., Chau, T.T.T.,
 Chevallier, F., Chini, L.P., Cronin, M., Currie, K.I., Decharme, B., Djeutchouang, L.M., Dou,
 X., Evans, W., Feely, R.A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor,
 L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R.A., Hurtt, G.C., Iida, Y., Ilyina, T., Luijkx,
 I.T., Jain, A., Jones, S.D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken,
 J.I., Körtzinger, A., Landschützer, P., Lauvset, S.K., Lefèvre, N., Lienert, S., Liu, J., Marland,
 G., McGuire, P.C., Melton, J.R., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S.-I., Niwa, Y., Ono,
 T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan,
 T.M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A.J., Sweeney, C., Tanhua, T.,
 Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G.R., Vuichard, N., Wada, C.,
 Wanninkhof, R., Watson, A.J., Willis, D., Wiltshire, A.J., Yuan, W., Yue, C., Yue, X., Zaehle,

S., Zeng, J. (2022). Global Carbon Budget 2021. *Earth System Science Data*, *14*(4), 1917–2005. DOI: 10.5194/essd-14-1917-2022

- Gelman, A. (1996). Bayesian model-building by pure thought: Some principles and examples. *Statistica Sinica*, *6*(1), 215–232.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC. ISBN: 1-58488-388-X
- Gelman, A., & Yao, Y. (2020). Holes in Bayesian statistics. *Journal of Physics G: Nuclear and Particle Physics*, 48(1), 014002. DOI: 10.1088/1361-6471/abc3a5
- Gogna, A., & Tayal, A. (2013). Metaheuristics: Review and application. *Journal of Experimental & Theoretical Artificial Intelligence*, *25*(4), 503–526. DOI: 10.1080/0952813X.2013.782347
- Goodfellow, I., Bengio, Y., & Courville, A. (2017). Deep learning (adaptive computation and machine learning series). *Cambridge Massachusetts*, 321–359. ISBN: 978-0-262-03561-3
- Guo, D., Westra, S., & Maier, H. R. (2016). An R package for modelling actual, potential and reference evapotranspiration. *Environmental Modelling & Software*, 78, 216–224. DOI: 10.1016/j.envsoft.2015.12.019
- Hararuk, O., Xia, J., & Luo, Y. (2014). Evaluation and improvement of a global land model against soil carbon data using a Bayesian Markov chain Monte Carlo method. *Journal of Geophysical Research: Biogeosciences*, *119*(3), 403–417. DOI: 10.1002/2013JG002535
- Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models–theory and application. *Ecology Letters*, *14*(8), 816–827. DOI: 10.1111/j.1461-0248.2011.01640.x
- Her, Y., & Chaubey, I. (2015). Impact of the numbers of observations and calibration parameters on equifinality, model performance, and output and parameter uncertainty. *Hydrological Processes*, 29(19), 4220–4237.DOI: 10.1002/hyp.10487

- Her, Y., & Seong, C. (2018). Responses of hydrological model equifinality, uncertainty, and performance to multi-objective parameter calibration. *Journal of Hydroinformatics*, *20*(4), 864–885. DOI: 10.2166/hydro.2018.108
- Hidy, D., Barcza, Z., Haszpra, L., Churkina, G., Pintér, K., & Nagy, Z. (2012). Development of the Biome-BGC model for simulation of managed herbaceous ecosystems. *Ecological Modelling*, 226, 99–119. DOI: 10.1016/j.ecolmodel.2011.11.008
- Hidy, D., Barcza, Z., Hollós, R., Dobor, L., Ács, T., Zacháry, D., Filep, T., Pásztor, L., Incze, D., & Dencső, M. (2022). Soil-related developments of the Biome-BGCMuSo v6. 2 terrestrial ecosystem model. *Geoscientific Model Development*, 15(5), 2157–2181. DOI: 10.5194/gmd-15-2157-2022
- Hidy, D., Barcza, Z., Hollós, R., Thornton, P., Running, S. W., & Fodor, N. (2021). User's Guide for Biome-BGC MuSo 6.2. http://nimbus.elte.hu/bbgc/files/Manual_BBGC_MuSo_v6.2.pdf
- Hidy, D., Barcza, Z., Marjanovic, H., Sever, M. Z. O., Dobor, L., Gelybó, G., Fodor, N., Pintér, K., Churkina, G., & Running, S. (2016). Terrestrial ecosystem process model Biome-BGCMuSo v4. 0: Summary of improvements and new modeling possibilities. *Geoscientific Model Development*, 9(12), 4405. DOI: 10.5194/gmd-9-4405-2016
- Hidy, D., Haszpra, L., Barcza, Z., Churkina, G., Trusilova, K., & Tomelleri, E. (2006). Bayesian calibration of the Biome-BGC C3 grass submodel. *Geophysical Research Abstracts*, *8*, 06831.
 SRef-ID: 1607-7962/gra/EGU06-A-06831
- Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science*, 3(2), 200–215. DOI: 10.1177/2515245919898657
- Hipp, R. D. (2020). SQLite (3.31.1). https://www.sqlite.org/index.html
- Hollós, R., Fodor, N., Merganičová, K., Hidy, D., Árendás, T., Grünwald, T., & Barcza, Z. (2022). Conditional interval reduction method: A possible new direction for the optimization of

process based models. *Environmental Modelling & Software*, *158*, 105556. DOI: 10.1016/j.envsoft.2022.105556

- Hollós, R., Kristóf, E., Fodor N., Hidy D., Horváth F., & Barcza Z. (2023). *RBBGCMuso: An R package to support the application of the Biome-BGCMuSo biogeochemical model* [Computer software]. https://github.com/hollorol/RBBGCMuso
- Hungerford, R.D., Nemani, R.R., Running, S.W., Coughlan, J.C., 1989. MTCLIM: a mountain microclimate simulation model. U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station. DOI: 10.2737/int-rp-414
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer. ISBN: 978-1461471370
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C., & Wheeler, T. R. (2017). Brief history of agricultural systems modeling. *Agricultural Systems*, 155, 240–254. DOI: 10.1016/j.agsy.2016.05.014
- Jönsson, A. M., Schroeder, L. M., Lagergren, F., Anderbrant, O., & Smith, B. (2012). Guess the impact of Ips typographus—An ecosystem modelling approach for simulating spruce bark beetle outbreaks. *Agricultural and Forest Meteorology*, *166–167*, 188–200. DOI: 10.1016/j.agrformet.2012.07.012
- Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., & Lichtenberger, J. (2018). Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agricultural and Forest Meteorology*, 260–261, 300–320. DOI: 10.1016/j.agrformet.2018.06.009
- Kern, A., Dobor, L., Hollós, R., Marjanović, H., Torma, CS.Z., Kis, A., Fodor, N., & Barcza, Z., (2024). Seamlessly combined historical and projected daily meteorological datasets for impact studies in Central Europe: The FORESEE v4.0 and the FORESEE-HUN v1.0. *Climate Services*, 33, 100443. DOI: 10.1016/j.cliser.2023.100443

- Khanfor, A., & Yang, Y. (2017). An Overview of Practical Impacts of Functional Programming. 2017
 24th Asia-Pacific Software Engineering Conference Workshops (APSECW), 50–54. DOI: 10.1109/APSECW.2017.27
- Kipling, R.P., Virkajärvi, P., Breitsameter, L., Curnel, Y., De Swaef, T., Gustavsson, A.-M., Hennart, S., Höglind, M., Järvenranta, K., Minet, J., Nendel, C., Persson, T., Picon-Cochard, C., Rolinski, S., Sandars, D.L., Scollan, N.D., Sebek, L., Seddaiu, G., Topp, C.F.E., Twardy, S., Van Middelkoop, J., Wu, L., Bellocchi, G. (2016). Key challenges and priorities for modelling European grasslands under climate change. *Science of The Total Environment*, *566–567*, 851–864. DOI: 10.1016/j.scitotenv.2016.05.144
- Krinner, G., Viovy, N., De Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P.,
 Sitch, S., & Prentice, I. C. (2005). A dynamic global vegetation model for studies of the
 coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, *19*(1),
 2003GB002199. DOI: 10.1029/2003GB002199
- LeBauer, D.S., Wang, D., Richter, K.T., Davidson, C.C., Dietze, & M.C., (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecological Monographs*, 83, 133–154. DOI: 10.1890/12-0137.1
- Lee, H., & Kim, S. (2016). Black-box classifier interpretation using decision tree and fuzzy logicbased classifier implementation. *International Journal of Fuzzy Logic and Intelligent Systems*, 16(1), 27–35. DOI: 10.5391/IJFIS.2016.16.1.27
- Lenton, T. M., Williamson, M. S., Edwards, N. R., Marsh, R., Price, A. R., Ridgwell, A. J., Shepherd,
 J. G., Cox, S. J., & The GENIE team. (2006). Millennial timescale carbon cycle and climate change in an efficient Earth system model. *Climate Dynamics*, *26*, 687–711. DOI: 10.1007/s00382-006-0109-9
- Li, C. (1996). The DNDC Model. In D. S. Powlson, P. Smith, & J. U. Smith (Eds.), *Evaluation of Soil Organic Matter Models* (pp. 263–267). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-61094-3_20

- Lovett, G. M., Cole, J. J., & Pace, M. L. (2006). Is Net Ecosystem Production Equal to Ecosystem Carbon Accumulation? *Ecosystems*, 9(1), 152–155. DOI: 10.1007/s10021-005-0036-3
- Ma, S., Churkina, G., Wieland, R., & Gessler, A. (2011). Optimization and evaluation of the ANTHRO-BGC model for winter crops in Europe. *Ecological Modelling*, *222*(20–22), 3662–3679. DOI: 10.1016/j.ecolmodel.2011.08.025
- McDermid, S.P., Ruane, A.C., Rosenzweig, C., Hudson, N.I., Morales, M.D., Agalawatte, P., Ahmad, S., Ahuja, L.R., Amien, I., Anapalli, S.S., Anothai, J., Asseng, S., Biggs, J., Bert, F., Bertuzzi, P., Bhatia, V.S., Bindi, M., Broad, I., Cammarano, D., Carretero, R., Chattha, A.A., Chung, U., Debats, S., Deligios, P., De Sanctis, G., Dhliwayo, T., Dumont, B., Estes, L., Ewert, F., Ferrise, R., Gaiser, T., Garcia, G., Gbegbelegbe, S., Geethalakshmi, V., Gerardeaux, E., Goldberg, R., Grant, B., Guevara, E., Hickman, J., Hoffmann, H., Huang, H., Hussain, J., Justino, F.B., Karunaratne, A.S., Koehler, A.-K., Kouakou, P.K., Kumar, S.N., Lakshmanan, A., Lieffering, M., Lin, X., Luo, Q., Magrin, G., Mancini, M., Marin, F.R., Marta, A.D., Masutomi, Y., Mavromatis, T., McLean, G., Meira, S., Mohanty, M., Moriondo, M., Nasim, W., Negm, L., Orlando, F., Orlandini, S., Ozturk, I., Soares Pinto, H.M., Podesta, G., Qi, Z., Ramarohetra, J., ur Rahman, M.H., Raynal, H., Rodriguez, G., Rötter, R., Sharda, V., Shuo, L., Smith, W., Snow, V., Soltani, A., Srinivas, K., Sultan, B., Swain, D.K., Tao, F., Tesfaye, K., Travasso, M.I., Trombi, G., Topaj, A., Vanuytrecht, E., Viscarra, F.E., Aftab Wajid, S., Wang, E., Wang, H., Wang, J., Wijekoon, E., Byun-Woo, L., Xiaoguang, Y., Young, B.H., Yun, J.I., Zhao, Z., Zubair, L. (2014). The AgMIP Coordinated Climate-Crop Modeling Project (C3MP): Methods and Protocols. In Handbook of Climate Change and Agroecosystems: Vol. 3 (pp. 191–220). Imperial College Press. DOI: 10.1142/9781783265640_0008
- Meersche, K. V. den, Soetaert, K., & Oevelen, D. V. (2009). xsample(): An R Function for Sampling
 Linear Inverse Problems. *Journal of Statistical Software*, 30, 1–15. DOI: 10.18637/jss.v030.c01

- Metivier, K. A., Pattey, E., & Grant, R. F. (2009). Using the ecosys mathematical model to simulate temporal variability of nitrous oxide emissions from a fertilized agricultural soil. *Soil Biology and Biochemistry*, *41*(12), 2370–2386. DOI: 10.1016/j.soilbio.2009.03.007
- Miyauchi, T., Machimura, T., & Saito, M. (2019). Estimating carbon fixation of plant organs for afforestation monitoring using a process-based ecosystem model and ecophysiological parameter optimization. *Ecology and Evolution*, *9*(14), 8025–8041. DOI: 10.1002/ece3.5328
- Moeller, C., Sauerborn, J., De Voil, P., Manschadi, A. M., Pala, M., & Meinke, H. (2014). Assessing the sustainability of wheat-based cropping systems using simulation modelling: Sustainability = 42? *Sustainability Science*, 9(1), 1–16. https://doi.org/10.1007/s11625-013-0228-2
- Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol'sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, *26*(12), 1515–1525. DOI: 10.1016/j.envsoft.2011.08.010
- Olesen, J. E., Trnka, M., Kersebaum, K. C., Skjelvåg, A. O., Seguin, B., Peltonen-Sainio, P., Rossi, F., Kozyra, J., & Micale, F. (2011). Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy*, *34*(2), 96–112. DOI: 10.1016/j.eja.2010.11.003
- Oleson, K. W., Niu, G.-Y., Yang, Z.-L., Lawrence, D. M., Thornton, P. E., Lawrence, P. J., Stöckli, R., Dickinson, R. E., Bonan, G. B., Levis, S., Dai, A., & Qian, T. (2008). Improvements to the Community Land Model and their impact on the hydrological cycle. *Journal of Geophysical Research: Biogeosciences*, 113, G01021, DOI: 10.1029/2007JG000563
- Parton, W.J. (1996). The CENTURY model. In: Powlson, D.S., Smith, P., Smith, J.U. (eds) Evaluation of Soil Organic Matter Models. NATO ASI Series, vol 38. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-61094-3_23
- Parton, W. J., Hartman, M., Ojima, D., & Schimel, D. (1998). DAYCENT and its land surface submodel: Description and testing. *Global and Planetary Change*, *19*(1–4), 35–48. DOI: 10.1016/S0921-8181(98)00040-X
- Parton, W. J., Schimel, D. S., Cole, C. V., & Ojima, D. S. (1987). Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands. *Soil Science Society of America Journal*, 51(5), 1173–1179. DOI: 10.2136/sssaj1987.03615995005100050015x
- Pásztor, L., Laborczi, A., Takács, K., Illés, G., Szabó, J., & Szatmári, G. (2020). Progress in the elaboration of GSM conform DSM products and their functional utilization in Hungary. *Geoderma Regional*, *21*, e00269. DOI: 10.1016/j.geodrs.2020.e00269
- Peng, C. (2000). From static biogeographical model to dynamic global vegetation model: A global perspective on modelling vegetation dynamics. *Ecological Modelling*, *135*(1), 33–54. DOI: 10.1016/S0304-3800(00)00348-3
- Pericchi, L. R., & Walley, P. (1991). Robust Bayesian Credible Intervals and Prior Ignorance. International Statistical Review / Revue Internationale de Statistique, 59(1), 1–23. DOI: 10.2307/1403571
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil*, *7*(1), 217–240. DOI: 10.5194/soil-7-217-2021
- Prescher, A.-K., Grünwald, T., & Bernhofer, C. (2010). Land use regulates carbon budgets in eastern Germany: From NEE to NBP. *Agricultural and Forest Meteorology*, *150*(7–8), 1016–1025. DOI: 10.1016/j.agrformet.2010.03.008
- Prihodko, L., Denning, A. S., Hanan, N. P., Baker, I., & Davis, K. (2008). Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. *Agricultural and Forest Meteorology*, 148(2), DOI: 268–287. 10.1016/j.agrformet.2007.08.006
- Pullens, J. W. M., Bagnara, M., Silveyra González, R., Gianelle, D., Sottocornola, M., Heijmans, M. M. P. D., Kiely, G., & Hartig, F. (2017). The NUCOMBog R package for simulating vegetation, water, carbon and nitrogen dynamics in peatlands. *Ecological Informatics*, *40*, 35–39. DOI: 10.1016/j.ecoinf.2017.05.001
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

- Raj, R., Hamm, N. A. S., van der Tol, C., & Stein, A. (2014). Variance-based sensitivity analysis of BIOME-BGC for gross and net primary production. *Ecological Modelling*, 292, 26–36. DOI: 10.1016/j.ecolmodel.2014.08.012
- Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., Zscheischler, J., Beer, C., Buchmann, N., & Frank, D. C. (2013). Climate extremes and the carbon cycle. *Nature*, 500(7462), 287–295. DOI: 10.1038/nature12350
- Ren, H., Zhang, L., Yan, M., Tian, X., & Zheng, X. (2022). Sensitivity analysis of Biome-BGCMuSo for gross and net primary productivity of typical forests in China. *Forest Ecosystems*, 9, 100011. DOI: 10.1016/j.fecs.2022.100011
- Rohmer, J., & Foerster, E. (2011). Global sensitivity analysis of large-scale numerical landslide models based on Gaussian-Process meta-modeling. *Computers & Geosciences*, *37*(7), 917–927. DOI: 10.1016/j.cageo.2011.02.020
- Roxburgh, S. H., Berry, S. L., Buckley, T. N., Barnes, B., & Roderick, M. L. (2005). What is NPP? Inconsistent accounting of respiratory fluxes in the definition of net primary production. *Functional Ecology*, *19*(3), 378–382. DOI: 10.1111/j.1365-2435.2005.00983.x
- Running, S. W., & Coughlan, J. C. (1988). A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modelling*, 42(2), 125–154. DOI: 10.1016/0304-3800(88)90112-3
- Sadegh, M., & Vrugt, J. A. (2013). Bridging the gap between GLUE and formal statistical approaches:
 Approximate Bayesian computation. *Hydrology and Earth System Sciences*, *17*(12), 4831–4850. DOI: 10.5194/hess-17-4831-2013
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. John Wiley & Sons. ISBN: 978-0-470-87094-5
- Sándor, R., Barcza, Z., Hidy, D., Lellei-Kovács, E., Ma, S., & Bellocchi, G. (2016). Modelling of grassland fluxes in Europe: Evaluation of two biogeochemical models. *Agriculture*, *Ecosystems & Environment*, *215*, 1–19. DOI: 10.1016/j.agee.2015.09.001

- Sexton, J., Everingham, Y., & Inman-Bamber, G. (2016). A theoretical and real world evaluation of two Bayesian techniques for the calibration of variety parameters in a sugarcane crop model. *Environmental Modelling and Software*, 83, 126–142. DOI: 10.1016/j.envsoft.2016.05.014
- Shevliakova, E., Stouffer, R. J., Malyshev, S., Krasting, J. P., Hurtt, G. C., & Pacala, S. W. (2013). Historical warming reduced due to enhanced land carbon uptake. *Proceedings of the National Academy of Sciences*, *110*(42), 16730–16735. DOI: 10.1073/pnas.1314047110
- Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., & Despouy, P. (2017). *plotly: Create Interactive Web Graphics via "plotly.js."* https://CRAN.Rproject.org/package=plotly
- Sitch, S., Prentice, I., Smith, B., Cramer, W., Kaplan, J., Lucht, W., Sykes, M., Thonicke, K., & Venevsky, S. (2000). *LPJ A coupled model of vegetation dynamics and the terrestrial carbon cycle*. Lund University, Lund,Sweden.
- Stedinger, J. R., Vogel, R. M., Lee, S. U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, 44. DOI: 10.1029/2008WR006822
- Steinau, S., Marrella, A., Andrews, K., Leotta, F., Mecella, M., & Reichert, M. (2019). DALEC: A framework for the systematic evaluation of data-centric approaches to process management software. *Software & Systems Modeling*, *18*, 2679–2716. DOI: 10.1007/s10270-018-0695-0
- Stöckle, C. O., Donatelli, M., & Nelson, R. (2003). CropSyst, a cropping systems simulation model. *European Journal of Agronomy*, *18*(3–4), 289–307. DOI: 10.1016/S1161-0301(02)00109-0
- Tange, O. (2015). *GNU Parallel 20150322 ('Hellwig')* [Computer software]. Zenodo. DOI: 10.5281/ZENODO.16303
- Tarantola, A., (2005). Inverse problem theory and methods for model parameter estimation. Society for Industrial and Applied Mathematics, Philadelphia, PA. ISBN: 978-0-89871-572-9

- Tatarinov, F. A., & Cienciala, E. (2006). Application of BIOME-BGC model to managed forests: 1.
 Sensitivity analysis. *Forest Ecology and Management*, 237(1), 267–279. DOI: 10.1016/j.foreco.2006.09.085
- Teuling, A. J., Seneviratne, S. I., Stöckli, R., Reichstein, M., Moors, E., Ciais, P., Luyssaert, S., Van Den Hurk, B., Ammann, C., & Bernhofer, C. (2010). Contrasting response of European forest and grassland energy exchange to heatwaves. *Nature Geoscience*, *3*(10), 722–727. DOI: 10.1038/ngeo950
- Therneau, T., Atkinson, B., Ripley, B. (producer of the initial R port & maintainer 1999-2017) (2022). *rpart: Recursive Partitioning and Regression Trees* (4.1.16) [Computer software]. https://CRAN.R-project.org/package=rpart
- Therond, O., Hengsdijk, H., Casellas, E., Wallach, D., Adam, M., Belhouchette, H., Oomen, R., Russell, G., Ewert, F., & Bergez, J.-E. (2011). Using a cropping system model at regional scale: Low-data approaches for crop management information and model calibration. *Agriculture, Ecosystems & Environment, 142*(1–2), 85–94. DOI: 10.1016/j.agee.2010.05.007
- Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo,
 L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., &
 Nikolova, S. (2007). OptIC project: An intercomparison of optimization techniques for
 parameter estimation in terrestrial biogeochemical models. *Journal of Geophysical Research: Biogeosciences*, *112*, *G02027*. DOI: 10.1029/2006jg000367
- Van Oijen, M., Rougier, J., & Smith, R. (2005). Bayesian calibration of process-based forest models:
 Bridging the gap between models and data. *Tree Physiology*, *25*(7), 915–927. DOI: 10.1093/treephys/25.7.915
- Van Vuuren, D. P., & Carter, T. R. (2014). Climate and socio-economic scenarios for climate change research and assessment: Reconciling the new with the old. *Climatic Change*, *122*, 415–429. DOI: 10.1007/s10584-013-0974-2

- Verbeeck, H., Samson, R., Verdonck, F., & Lemeur, R. (2006). Parameter sensitivity and uncertainty of the forest carbon flux model FORUG: A Monte Carlo analysis. *Tree Physiology*, *26*(6), 807–817. DOI: 10.1093/treephys/26.6.807
- Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hiremath, S., Hoek, S., Horan, H., Hoogenboom, G., Huang, M., Jabloun, M., Jansson, P.-E., Jing, Q., Justes, E., Kersebaum, K.C., Klosterhalfen, A., Launay, M., Lewan, E., Luo, Q., Maestrini, B., Mielenz, H., Moriondo, M., Nariman Zadeh, H., Padovan, G., Olesen, J.E., Poyda, A., Priesack, E., Pullens, J.W.M., Qian, B., Schütze, N., Shelia, V., Souissi, A., Specka, X., Srivastava, A.K., Stella, T., Streck, T., Trombi, G., Wallor, E., Wang, J., Weber, T.K.D., Weihermüller, L., de Wit, A., Wöhling, T., Xiao, L., Zhao, C., Zhu, Y., Seidel, S.J. (2021). The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling & Software*, *145*, 105206. DOI: 10.1016/j.envsoft.2021.105206
- Webber, H., Ewert, F., Olesen, J.E., Müller, C., Fronzek, S., Ruane, A.C., Bourgault, M., Martre, P., Ababaei, B., Bindi, M., Ferrise, R., Finger, R., Fodor, N., Gabaldón-Leal, C., Gaiser, T., Jabloun, M., Kersebaum, K.-C., Lizaso, J.I., Lorite, I.J., Manceau, L., Moriondo, M., Nendel, C., Rodríguez, A., Ruiz-Ramos, M., Semenov, M.A., Siebert, S., Stella, T., Stratonovitch, P., Trombi, G., Wallach, D. (2018). Diverging importance of drought stress for maize and winter wheat in Europe. *Nature Communications*, *9*, 4249. DOI: 10.1038/s41467-018-06525-2
- White, M. A., Thornton, P. E., Running, S. W., & Nemani, R. R. (2000). Parameterization and sensitivity analysis of the BIOME–BGC terrestrial ecosystem model: Net primary production controls. *Earth Interactions*, 4(3), 1–85. DOI: 10.1175/1087-3562(2000)004<0003:PASAOT>2.0.CO;2
- Wickham, H., Danenberg, P., Csárdi, G., & Eugster, M. (2022). *roxygen2: In-Line Documentation for R*. https://CRAN.R-project.org/package=roxygen2

- Wickham, H., Hester, J., Chang, W., & Bryan, J. (2022). *devtools: Tools to Make Developing R Packages Easier*. https://CRAN.R-project.org/package=devtools
- Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., & Wang, Y.-P. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, 6(7), 1341–1359. DOI: 10.5194/bg-6-1341-2009
- Wöhling, T., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H.-D., Högy, P., Cuntz, M., Attinger, S., Wulfmeyer, V., & Streck, T. (2013). Multiresponse, multiobjective calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and CLM3. 5. *Water Resources Research*, 49(12), 8200–8221. DOI: https://doi.org/10.1002/2013WR014536
- Xiong, W., Holman, I., Conway, D., Lin, E., & Li, Y. (2008). A crop model cross calibration for use in regional climate impacts studies. *Ecological Modelling*, *213*(3–4), 365–380. DOI: 10.1016/j.ecolmodel.2008.01.005
- Zabinsky, Z. B., & Smith, R. L. (2013). Hit-and-Run Methods. In S. I. Gass & M. C. Fu (Eds.), *Encyclopedia of Operations Research and Management Science* (pp. 721–729). Springer US. DOI: 10.1007/978-1-4419-1153-7_1145
- Zelenák, A., Szabó, A., Nagy, J., & Nyéki, A. (2022). Using the Ceres-Maize model to simulate crop yield in a long-term field experiment in Hungary. *Agronomy*, *12*(4), 785. DOI: 10.3390/agronomy12040785
- Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., & Zhang, X. (2018). Future climate risk from compound events. *Nature Climate Change*, *8*(6), 469–477. DOI: 10.1038/s41558-018-0156-3

MELLÉKLET

A Biome-BGCMuSo 6.3 teljes ökofiziológiai paraméterezése kukoricára

Az M1. táblázat a teljes a priori paraméterezést tartalmazza, amely alapján a dolgozatban bemutatott CIRM alapú eljárást demonstráltam. Az M2. táblázat a teljes a poszteriori parametrizációt közli.

M1. táblázat. A priori ökofiziológiai paraméterezés kukoricára. Néhány változó nem releváns a lágyszárú növényre vonatkozó szimuláció miatt.

ECOPHYS FILE	- maize						
FLAGS							
0	(flag)	biome type flag (1 = WOODY 0 = NON-WOODY)					
0	(flag)	woody type flag (1 = EVERGREEN 0 = DECIDUOUS)					
0	(flag)	photosynthesis type flag (1 = C3 PSN 0 = C4 PSN)					
PLANT FUNCTIO	NING PARAMETERS						
Ō	(yday)	yearday to start new growth					
364	(yday)	yearday to end litterfall					
0.4	(prop.)	transfer growth period as fraction of growing season					
0	(prop.)	litterfall as fraction of growing season					
8	(Celsius)	base temperature					
-9999	(Celsius)	minimum temperature for growth displayed on current day					
-9999	(Celsius)	optimali temperature for growth displayed on current day					
-9999	(Celsius)	maxmimum temperature for growth displayed on current day					
-9999	(Celsius)	minimum temperature for carbon assimilation displayed on current day					
-9999	(Celsius)	optimal1 temperature for carbon assimilation displayed on current day					
-9999	(Celsius)	optimal2 temperature for carbon assimilation displayed on current day					
-9999	(Celsius)	maxmimum temperature for carbon assimilation displayed on current day					
1	(1/yr)	annual leaf and fine root turnover fraction					
0	(1/yr)	annual live wood turnover fraction					
0	(1/yr)	annual fire mortality fraction					
10	(1/vegper)	Whole-plant mortality fraction in vegetation period					
10	(kgC/kgN) (kgC/kgN)	C:N of loaf littor after retranslocation					
58	(kgC/kgN)	C.N of fine roots					
50	(kgC/kgN)	C:N of fruit					
71	(kgC/kgN)	C:N of soft stem					
0	(kgC/kgN)	C:N of live wood					
0	(kgC/kgN)	C:N of dead wood					
0.4	(kgC/kgDM)	dry matter carbon content of leaves					
0.4	(kgC/kgDM)	dry matter carbon content of leaf litter					
0.4	(kgC/kgDM)	dry matter carbon content of fine roots					
0.4	(kgC/kgDM)	dry matter carbon content of iruit					
0.4	(kgC/kgDM) (kgC/kgDM)	dry matter carbon content of live wood					
0.4	(kgC/kgDM)	dry matter carbon content of dead wood					
0.68	(DIM)	leaf litter labile proportion					
0.23	(DIM)	leaf litter cellulose proportion					
0.34	(DIM)	fine root labile proportion					
0.44	(DIM)	fine root cellulose proportion					
0.68	(DIM)	fruit litter labile proportion					
0.23	(DIM)	soft stom litter labile propertion					
0.23	(DIM)	soft stem litter cellulose proportion					
0	(DIM)	dead wood cellulose proportion					
0.002	(1/LAI/d)	canopy water interception coefficient					
0.75	(DIM)	canopy light extinction coefficient					
2	(g/MJ)	potential radiation use efficiency					
0.781	(DIM)	radiation parameter1					
13.596	(DIM)	radiation parameter2					
2	(DIM)	all-Slaed to projected lear area ratio					
0.1	(DIM)	fraction of leaf N in Rubisco					
0.03	(DIM)	fraction of leaf N in PEP Carboxvlase					
0.006	(m/s)	maximum stomatal conductance (projected area basis)					
0.00006	(m/s)	cuticular conductance (projected area basis)					
0.0025	(m/s)	boundary layer conductance (projected area basis)					
2	(m)	maximum height of plant					
0.16	(KgC)	stem weight corresponding to maximum height					
1.8	(uiiiiiess)	prant nerght runction snape parameter (stope)					
±.5 3 67	(DTM)	root distribution parameter					
0.1	(kgC)	root weight corresponding to max root depth					
1.3	(dimless)	root depth function shape parameter (slope)					
1000	(m/kg)	root weight to rooth length conversion factor					

0.3		(prop.)		growth resp pe	er unit of	C grown			
0.218		(kgC/kgN/d)		maintenance re	espiration	in kgC/day per	kg of tissue N		
0.1		(DIM)		theoretical max. prop. of non-structural and structural carbohydrate					
0.3		(DIM)		prop. of non-structural carbohydrates available for maintanance respiration					
0.0005		(kgN/m2/yr)		symbiotic+asyr	nbiotic N	fixation			
0		(day)		time delay for temperature in photosynthesis acclimation					
CROP SPE	CIFIC	PARAMETERS							
1		(DIM)		number of pher	nophase of	germination			
2		(DIM)		number of pher	nophase of	emergence			
0.3		(prop.)		critical VWCra	atio (prop	. to FC-WP) in	germination		
0		(DIM)		number of pher	nophase of	photoperiodic	slowing effect		
16		(hour)		critical photo	slow davl	ength			
0.005		(DTM)		slope of relat	tive photo	slow developmer	t rate		
0		(DTM)		number of pher	nophase of	vernalization			
0		(Celsius)		critical verna	lization	temperature 1			
5		(Colsius)		critical verna	lization	temperature 2			
8		(Colsius)		critical verna	lization	temperature 3			
15		(Colsius)		critical vorm	lization	tomporaturo 4			
1 0 0 1		(DTM)		clicical velua	iiizacion	lightion develo	mont vata		
50		(DIM) (D)		stope of retai	ligation	dava (in warmal	ization development rate)		
50		(II) (DTM)		required verna	arrzacion	uays (in vernai	ización development face)		
5		(DIM)		number of itov	vering phe	nopnase			
30		(Cersius)		Critical llowe	ering neau	stress tempera	iture i		
40		(Celsius)		critical flowe	ering heat	stress tempera	ture 2		
0.2		(prop.)		theoretical ma	aximum of	flowering therm	al stress mortality parameter		
STRESS A	ND SEN	ESCENCE PARAM	IETERS						
0.6		(prop)		VWC ratio to d	calc. soil	moisture limit	I (prop. to FC-WP)		
0.95		(prop)		VWC ratio to d	calc. soil	moisture limit	2 (prop. to SAT-FC)		
0.4		(prop)		minimum of so:	il moistur	e limit2 multip	licator		
3000		(Pa)		vapor pressure	e deficit:	start of condu	ictance reduction		
5200		(Pa)		vapor pressure	e deficit:	complete condu	ctance reduction		
0.017		(prop.)		max. senescend	ce mortali	ty coefficient	of aboveground plant material		
0.006		(prop.)		max. senescend	ce mortali	ty coefficient	of belowground plant material		
0		(prop.)		max. senescence mortality coefficient of non-structured plant material					
35		(Celsius)		lower limit extreme high temperature effect on senescence mortality					
40		(Celsius)		upper limit ex	ktreme hig	h temperature e	ffect on senescence mortality		
0.001		(prop.)		turnover rate of wilted standing biomass to litter					
0.03		(prop.)		turnover rate	of non-wo	ody cut-down bi	omass to litter		
0.01		(prop.)		turnover rate	of woody	cut-down biomas	s to litter		
15		(nday)		drought tolera	ance param	eter (critical	value of DSWS)		
0		(dimless)		effect of soil	lstress on	photosynthesis			
GROWING	SEASON	I PARAMETERS							
5		(kg/m2)		crit. amount o	of snow li	miting photosyr	thesis		
20		(Celsius)		limit1 (under:	full cons	trained) of HEA	TSUM index		
60		(Celsius)		limit2 (above:	unconstra	ined) of HEATSU	M index		
0		(Celsius)		limit1 (under:	full cons	trained) of TMI	N index		
5		(Celsius)		limit2 (above:	unconstra	ined) of TMIN i	ndex		
6000		(Pa)	Pa) limit1 (above:full constrained) of VPD index						
1000		(Pa) limit2 (under:unconstrained) of VPD index							
0		(s)	s) limit1 (under:full constrained) of DAYLENGTH index						
0		(s)		limit2 (above:	unconstra	ined) of DAYLEN	IGTH index		
10		(dav)		moving window	length				
0.1		(dimless)		GST limit1 (a	reater tha	t limit -> star	t of vegper)		
0.01		(dimless)		GST limit2 (le	ess that l	imit -> end of	vegper)		
ALLOCATI	ION PAF	AMETERS (7 ph	nenoloc	rical phases)					
phase1	phase2	2 phase3	phase	4 phase5	phase6	phase7	(text) name of the phenophase		
10	50	240	240	75	1000	10000	(Celsius) length of phenophase (GDD)		
0.5	0.50	0.40	0.3	0.05	0.0	0.0	(ratio) leaf ALLOCATION		
0.5	0.50	0.30	0.40	0.20	0.15	0.0	(ratio) fine root ALLOCATION		
0.0	0.00	0.0	0.60	0.05	0.7	0.0	(ratio) fruit ALLOCATION		
0	0 00	0.30	0 30	0.70	0.15	0.0	(ratio) soft stem ALLOCATION		
0	0	0.50	0	0.70	0	0	(ratio) live woody stem ALLOCATION		
õ	ő	0	Ő	0	Ő	0	(ratio) dead woody stem AlloCATION		
õ	ő	0	ő	0	0	0	(ratio) live coarse root ALLOCATION		
õ	ő	0	Ő	0	Ő	0	(ratio) dead coarse root ALLOCATION		
60	50	50	45	35	25	30	(m2/kgC) canony average specific leaf area		
1 0	1 0	1 0	-⊒J 1 ∩	1 0	1 0	1 0	(mz/kyc) canopy average specific fed afed		
1000	1500	1500	1000	1000	1000	1000	(Celeius) maximal lifetime of plant ticeus		
T 0 0 0	1000	T 3 0 0	T000	1000	T 0 0 0	1000	(certing) maximar filectime of plant tissue		

M2. táblázat. A poszteriori ökofiziológiai paraméterezés kukoricára.

ECOPIIIS FILE									
FLAGS									
0	(flag)	biome type flag (1 = WOODY 0 = NON-WOODY)							
0	(flag)	woody type flag (1 = EVERGREEN 0 = DECIDUOUS)							
0	(flag)	photosyn. type flag (1 = C3 PSN 0 = C4 PSN)							
PLANT FUNCTIONING PARAMETERS									
0	(yday)	yearday to start new growth							
364	(yday)	yearday to end litterfall							
0.4	(prop.)	transfer growth period as fraction of growing season							
0	(prop.)	litterfall as fraction of growing season							
8	(Celsius)	base temperature							
-9999	(Celsius)	minimum temperature for growth displayed on current day							
-9999	(Celsius)	optimall temperature for growth displayed on current day							
-9999	(Celsius)	optimal2 temperature for growth displayed on current day							
-9999	(Celsius)	maxmimum temperature for growth displayed on current day							
-9999	(Celsius)	minimum temperature for carbon assimilation displayed on current day							
-9999	(Celsius)	optimal1 temperature for carbon assimilation displayed on current day							
-9999	(Celsius)	optimal2 temperature for carbon assimilation displayed on current day							
-9999	(Celsius)	maxmimum temperature for carbon assimilation displayed on current day							
1	(1/yr)	annual leaf and fine root turnover fraction							
0	(1/yr)	annual live wood turnover fraction							
0	(1/yr)	annual fire mortality fraction							
0.001	(1/vegper)	whole-plant mortality fraction in vegetation period							

18 65		(kgC/kgN) (kgC/kgN)	C:N	of leaves	s litter. aft	er retranslocation	
58		(kgC/kgN)	C:N	of fine r	roots	er recransiocación	
50 71		kgC/kgN) (kgC/kgN)	C:N C:N	of fruit of soft s	stem		
0		(kgC/kgN)	C:N	of live w	lood		
0.4		(kgC/kgN) (kgC/kgDM)	C:N drv	of dead w matter ca	vood arbon conte	nt of leaves	
0.4		(kgC/kgDM)	dry	matter ca	arbon conte	nt of leaf litter	
0.4		(kgC/kgDM) * (kgC/kgDM)	dry dry	matter ca matter ca	arbon conte arbon conte	nt of fine roots nt of fruit	
0.4		(kgC/kgDM)	dry	matter ca	arbon conte	nt of soft stem	
0.4		* (kgC/kgDM) * (kgC/kgDM)	dry dry	matter ca matter ca	arbon conte arbon conte	nt of live wood nt of dead wood	
0.68		(DIM)	leaf	litter l	labile prop	ortion	
0.23		(DIM) (DIM)	lea: fine	: litter d e root lab	cellulose p pile propor	roportion tion	
0.44		(DIM)	fine	root cel	llulose pro	portion	
0.68		* (DIM) * (DIM)	frui	t litter. t litter	cellulose	portion proportion	
0.68		(DIM)	soft	stem lit	ter labile	proportion	
0.23		(DIM) * (DIM)	dead	l wood cel	llulose pro	portion	
0.002		(1/LAI/d)	cano	py water	intercepti	on coefficient	
2		(g/MJ)	pote	ential rac	diation use	efficiency	
0.781		(DIM)	radi	ation par	ameter1		
2		(DIM) (DIM)	all-	sided to	projected	leaf area ratio	
2		(DIM)	rati	o of shad	ded SLA:sun	lit SLA	
0.00929		(DIM)	frac	tion of 1	leaf N in R	EP Carboxylase	
0.006		(m/s) (m/s)	maxi	mum stoma	atal conduc	tance (projected are	ea basis)
0.0025		(m/s)	bour	dary laye	er conducta	nce (projected area	basis)
2		(m) (lrac)	maxi	mum heigh	nt of plant	ng to mavimum boight	
0.10		(dimless)	plar	it height (function s	hape parameter (slop	pe)
1.8		(m) (DTM)	maxi	.mum depth	n of rootin	g zone	
0.092		(kgC)	root	: weight c	correspondi	eter ng to max root depth	1
0.758		(dimless)	root	depth fu	nction sha	pe parameter (slope)	or
0.3		(prop.)	grov	th resp p	per unit of	C grown	.01
0.218		(kgC/kgN/d)	mair thor	tenance r	espiration	in kgC/day per kg c	of tissue N
0.3		(DIM)	prop	. of non-	structural	carbohydrates avail	Lable for maintanance respiration
0.0005		(kgN/m2/yr) (day)	symk time	iotic+asy delav fo	mbiotic fi r temperat	xation of N ure in photosynthesi	s acclimation
CROP SPI	ECIFIC	PARAMETERS (DIM)	numk	er of phe	enophase of	germination	
2		(DIM)	numk	er of phe	enophase of	emergence	
0.3		(prop.) (DIM)	crit numb	er of phe	atio (prop enophase of	 to FC-WP) in germi photoperiodic slowi 	ination ing effect
16		(hour)	crit	ical phot	oslow dayl	ength	-
0.005		(DIM) (DIM)	numk	e or rela er of phe	ntive photo enophase of	slow development rat vernalization	.e
0		(Celsius)	crit	ical verr	alization	temperature 1	
8		(Celsius)	crit	ical vern	alization	temperature 3	
15		(Celsius)	crit	ical verr	alization	temperature 4 lization development	rate
50		(n)	requ	ired verr	alization	days (in vernalizati	ion development rate)
5 30		(DIM) (Celsius)	numb	er of flo	wering phe vering heat	nophase stress temperature	1
40		(Celsius)	crit	ical flow	vering heat	stress temperature	2
0.2		(prop.)	theo	retical m	naximum of	flowering thermal st	ress mortality parameter
STRESS A	AND SEN	ESCENCE PARAMETH	RS				
0.6		(prop)	VWC VWC	ratio to	calc. soil	moisture limit 1 (p	prop. to FC-WP) prop. to SAT-FC)
0.4		(prop)	mini	mum of so	oil moistur	e limit2 multiplicat	tor
3000 5200		(Pa) (Pa)	vapo vapo	or pressur or pressur	re deficit: re deficit:	start of conductanc complete conductanc	ce reduction ce reduction
0.01469		(prop.)	max.	senescer	nce mortali	ty coefficient of ab	ooveground plant material
0.006		(prop.)	max. max.	senescer	ice mortali ice mortali	ty coefficient of be ty coefficient of no	elowground plant material on-structured plant material
35		(Celsius)	lowe	er limit ∈	extreme hig	h temperature effect	on senescence mortality
40 0.001		(prop.)	uppe turr	er limit e Nover rate	extreme nig. e of wilted	n temperature effect standing biomass to	on senescence mortality
0.03		(prop.)	turr	over rate	e of non-wo	ody cut-down biomass	s to litter
15		(prop.) (nday)	drou	iover rate ight toler	ance param	cut-down biomass to eter (critical value	of DSWS)
0.1887		(dimless)	effe	ect of soi	llstress on	photosynthesis	
GROWING	SEASON	PARAMETERS					
5 20		(kg/m2) (Celsius)	crit limi	amount	of snow lin	miting photosyn.	index
60		(Celsius)	limi	.t2 (above	:unconstra	ined) of HEATSUM ind	lex
0 5		(Celsius) (Celsius)	limi limi	tl (under t2 (above	:full cons	trained) of TMIN ind ined) of TMIN index	lex
6000		(Pa)	limi	.t1 (above	:full cons	trained) of VPD inde	×
1000 0		(Pa) (s)	limi limi	.t2 (under .t1 (under	:unconstra	ined) of VPD index trained) of DAYLENGT	TH index
0		(s)	limi	t2 (above	:unconstra	ined) of DAYLENGTH i	index
10 0.1		(day) (dimless)	movi GSI	ng window. limit1 (c	v length greater tha	t limit -> start of	vegper)
0.01		(dimless)	GSI	limit2 (1	less that 1	imit -> end of vegpe	er)
ALLOCAT	ION PAR	AMETERS (7 pheno	logica	l phases)			
phase1	phase2	2 phase3	ase4	phase5	phase6	phase7	(text) name of the phenophase
0.5	0.50	0.417 0.	273	0.05	0.0	0.0	(ratio) leaf ALLOCATION
0.5	0.50 0.00	0.339 0. 0.0 0	334 000	0.20 0.05	0.095 0.657	0.0 0.0	(ratio) fine root ALLOCATION (ratio) fruit ALLOCATION
0.	0.00	0.244 0	393	0.70	0.248	0.0	(ratio) soft stem ALLOCATION

0	0	0	0	0	0	Ō	(ratio) live woody stem ALLOCATION
0	0	0	0	0	0	0	(ratio) dead woody stem ALLOCATION
0	0	0	0	0	0	0	(ratio) live coarse root ALLOCATION
0	0	0	0	0	0	0	(ratio) dead coarse root ALLOCATION
60	50	50	45	35	25	30	(m2/kgC) canopy average specific leaf area
1.0	1.0	1.0	1.0	1.0	1.0	1.0	(prop.) current growth proportion
1000	1000	930.7	965.8	1000	800	700	(Celsius) maximal lifetime of plant tissue

BIOGEOKÉMIAI MODELLRE ÉPÜLŐ SZIMULÁCIÓS KERETRENDSZER FEJLESZTÉSE ÉS ALKALMAZÁSA

Hollós Roland

Témavezetők: Barcza Zoltán, Fodor Nándor

Doktori munkám keretében a hazai fejlesztésű Biome-BGCMuSo folyamatorientált biogeokémiai modell alkalmazását támogató keretrendszereket fejlesztettem ki és alkalmaztam döntéshozói kontextusban. A fejlesztések vezérlőelve az volt, hogy a modell-adat fúzió minél több komponensére nyújtsak konkrét, a gyakorlatban is alkalmazható megoldást. Minden általam készített szoftver nyílt forráskódú (GPL-2 licensz), és GitHub-on szabadon elérhető.

A modell alkalmazását támogató, saját fejlesztésű RBBGCMuso keretrendszer elsősorban a kutatói közösséget célozza meg. A Biome-BGCMuSo modellcsalád kontextusában nem létezik hasonló fejlesztés a tudományos közösségben. Az RBBGCMuso részeként egy új eljárást dolgoztam ki a folyamatorientált modellek optimalizálására, amely túlmutat a hagyományos valószínűségi módszereken. Az általam javasolt ún. CIRM módszer leginkább a hagyományos valószínűségi módszerek és egy jól értelmezhető gépi tanulási módszer kombinációjaként írható le. Legjobb tudásom szerint a szakirodalomban nem található hasonló megközelítés. A munkám során pontszerű állomási adatok felhasználásával optimalizáltam a Biome-BGCMuSo-t a CIRM segítségével.

A saját fejlesztésű AgroMo rendszer összegzi a munkámat. Ebben letisztult formában, grafikus felhasználói felület mögé rejtve minden bemutatott komponens szerepel: a Biome-BGCMuSo, az RBBGCMuso, a FORESEE klímaadatbázis, a DoSoReMI talajadatbázis, az SQL alapú természetes query-nyelv, az ún. storyline funkcionalitás és a párhuzamosítás. Az AgroMo letölthető a nyílt elérésű GitHub felületről.

A dolgozat zárásaként bemutatott esettanulmány a jelenleg rendelkezésre álló legkorszerűbb alapadatokat használja, és reprodukálható módon vetíti előre a kukorica terméshozamát az RCP4.5, illetve 8.5 forgatókönyvek alapján az AgroMo felhasználásával. Mivel azzal a feltételezéssel éltünk, hogy az emberi beavatkozás jellege (vagyis az agrotechnika) nem változik, ezért az eredmények a változó időjárás és a növekvő légköri CO₂ koncentráció együttesen kifejtett hatását jelenítik meg. Az eredmények szerint a terméshozam csökkenése várható. Az RCP4.5 és 8.5 trajektóriák eltérnek, és nem intuitív időbeli alakulást mutatnak. Az eredmények rámutatnak további forgatókönyvek konstruálásának fontosságára, ahol klímaadaptív mezőgazdasági művelési módok bevezetését is számításba kell venni, mint pl. az öntözés, a vetési idő módosulása, és szárazságtűrő hibridek bevezetése.

DEVELOPMENT OF SIMULATION FRAMEWORKS TO SUPPORT BIOGEOCHEMICAL MODEL APPLICATION

Roland HOLLÓS

Supervisors: Zoltán BARCZA, Nándor FODOR

Biogeochemical models such as Biome-BGCMuSo are widely used to quantify plant growth, crop yield and the greenhouse gas budget of ecosystems. However, the high complexity of such models can hinder model use even at point scale. Moreover, these models need optimization prior to application that is challenging due to the high number of adjustable parameters.

The dissertation focuses on the development of a framework to support various elements of model application including input data construction, sensitivity analysis, spatial application, and visualization. Additionally, special emphasis was put on improving model optimization that helps estimating unknown parameters and thus reduces the uncertainty of the simulations.

One major achievement of the research is the construction of an application programming interface for Biome-BGCMuSo. This interface is called RBBGCMuso written in R language, and it includes the most common scientific workflows related to the model application. Currently there is no comparable development in the scientific community for the Biome-BGCMuSo model family. RBGCMuso is open source and is disseminated at GitHub to fulfill the requirements of the scientific community.

I also developed a novel procedure called CIRM that can be used to effectively optimize models. This method minimizes the uncertainty of model parameters. CIRM combines the traditional probabilistic methods with a well-interpretable machine learning method. To my best knowledge, there is no similar approach in scientific literature. I demonstrated the applicability of CIRM in a low-data environment to simulate crop yield. The model was then successfully validated at county (NUTS3) level.

Another milestone of my research is the so-called AgroMo system. This streamlined, graphical user interface based software summarizes the entire modelling workflow. It is built by integrating Biome-BGCMuSo, RBBGCMuSo, the FORESEE climate database, the DoSoReMI soil database, and an SQL-based novel query meta-language, also facilitating parallelization. It is open source software, and can be downloaded from GitHub.

Using AgroMo I carried out an impact study to quantify climate change effect on maize production in Hungary. The projections suggest that yield loss is likely in the future if human interventions remain unchanged. The results highlight the importance of constructing further scenarios where climate-adaptive agricultural practices are considered.