

FORTANNS

manuál

Vojtěch Havlíček
havlicekv@fzp.czu.cz

2. března 2010

Obsah

1	Úvod	3
2	Instalace, spuštění, vstupy	3
2.1	Instalace	3
2.2	Vstupní data	3
2.3	Spuštění	4
2.4	Načtení vstupních dat	4
2.5	Volba typu výpočtu	6
3	Normální výpočet	6
3.1	Nastavení fáze výpočtu (1. oblast)	7
3.2	Volba parametrů sítě (2. oblast)	8
3.3	Volba parametrů vstupů (3. oblast)	9
3.4	Volba vnitřní architektury sítě a sledování výstupů (4. oblast)	9
3.5	Pokročilé nastavení (5. oblast)	10
3.6	Potvrzení nastavení (6. oblast)	11
4	Optimalizace	11
5	Start výpočtu	12
6	Výstupy	13
6.1	Výstupy kalibrace	13
6.2	Výstupy testování a validace	13
6.3	Výstupy optimalizace	13
7	Příklady	14
7.1	Predikce odtoku	14
7.2	Predikce početních stavů rysa kanadského	15

1 Úvod

Program FORTANNS je software určený k modelování časových řad. Kód programu má 2960 řádek a je napsán v programovacím jazyku Fortran 95. Program je možné používat v operačním systému Linux. Software FORTANNS je vydán jako open source, což znamená, že je k dispozici společně se svými zdrojovými kódy a potenciální uživatel má možnost kód upravovat.

Jednotlivé verze programu je možné stáhnout na stránkách Katedry vodního hospodářství a environmentálního modelování (<http://www.kvhem.cz/vyzkum/software/>). Tento manuál je pro verzi FORTANNS 2009.

Název programu FORTANNS je zkratkou z anglického 'FORecasting Time series Artificial Neural Network System'. Program je určen, jak bylo zmíněno výše pro modelování a zejména následnou predikci časových řad různých veličin. Výpočetním jádrem programu je model umělých neuronových sítí (NS) fungující ve spolupráci s evolučními algoritmy pro optimalizaci parametrů neuronové sítě. Neuronová síť použitá v programu je vícevrstevný perceptron se zpětnou propagací chyby v síti. Ovládání programu nevyžaduje detailní znalost implementovaných metod, avšak obecné povědomí o těchto metodách je předpokladem efektivního použití programu FORTANNS. Tento manuál není věnován popisu modelu umělých neuronových sítí ani popisu evolučních algoritmů a proto je dále uvedena literatura zabývající se touto problematikou, ve které si může zájemce dohledat a osvěžit znalosti o výše zmíněných metodách. Pro seznámení se s umělými neuronovými sítěmi jsou vhodné publikace (Šíma, Neruda 1996), (Fausett 1994). Evoluční a genetické algoritmy jsou dobře popsány a vysvětleny v (Onwubolu, Babu 2004) nebo v tuzemské literatuře (Zelinka a kol. 2008), (Hynek J. 2008). Výše uvedená literatura je jen zlomkem možných zdrojů, které jsou k dané problematice v současnosti k dispozici.

Pohyb v oknech programu je možný myší + levé tlačítko na kliknutí nebo pomocí kláves Tab, šipek a mezerníku pro potvrzení (případně klávesy Enter).

Názvy složek v tomto manuálu jsou tučně a kurzívou, názvy souborů tučně, akce prováděné v oknech programu jsou kurzívou, popisy v oknech programu jsou v uvozovkách.

2 Instalace, spuštění, vstupy

2.1 Instalace

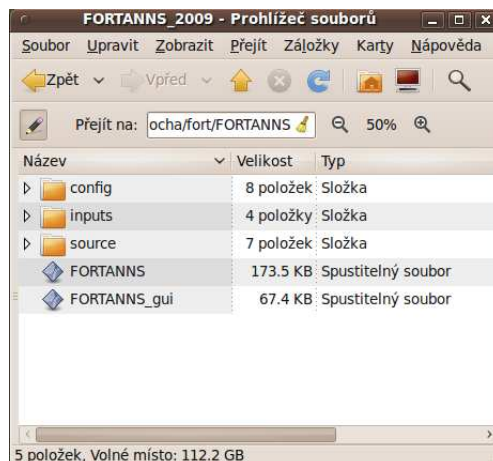
Program FORTANNS je ke stažení na webových stránkách KVHEM: <http://www.kvhem.cz/vyzkum/software/>. Po stažení a dekompresi souboru FORTANNS_2009.zip je možné začít hned pracovat - program nevyžaduje instalaci. Po úspěšném rozbalení archivu by měl adresář **FORTANNS_2009** obsahovat 3 složky (**config**, **source**, **inputs**) a 2 spustitelné soubory (**FORTANNS** a **FORTANNS_gui**), viz obr. 1.

Pro vlastní práci jsou důležité spustitelné soubory **FORTANNS** a **FORTANNS_gui** a složka **inputs**. Ve složce **source** jsou uloženy zdrojové kódy programu. Ve složce **config** se nachází konfigurační soubory programu, které je možné v určitých situacích editovat.

2.2 Vstupní data

Před vlastním spuštěním programu je dobré mít datové soubory, které budou určeny pro modelování, nahrány ve složce **inputs**. Tato složka je určena pro uložení souborů s daty, která chceme využít pro modelování.

Datové soubory pro práci programu musí být ve formě textového souboru. Data v souboru musí být v jednotlivých sloupcích, pro každou veličinu jeden sloupec. Oddělovač sloupců může být tabelátor nebo jedna či více mezer. Desetiný oddělovač musí být tečka. Po posledním řádku dat nenásleduje prázdný řádek.



Obrázek 1: Program FORTANNS po rozbalení

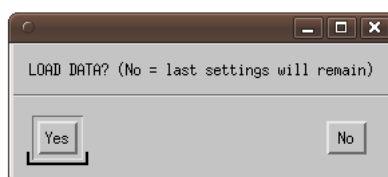
2.3 Spuštění

Dále popsany postup práce s programem FORTANNS je věnován ovládání pomocí GUI (Graphical User Interface).

Program se spouští z terminálu. Po zadání cesty do adresáře FORTANNS_2009 slouží pro spuštění programu příkaz `./FORTANNS_gui`. Tímto příkazem se nashoduje uživatelské grafické rozhraní, které slouží k ovládání programu.

2.4 Načtení vstupních dat

První dialogové okno, které se objeví na obrazovce (viz obr. 2) je dotaz na načítání datových souborů. Volit je možné varianty *Yes* nebo *No*. Při prvním použití programu je nutné navolit vstupní datové soubory, čili použít volbu *Yes*! V případě volby *No* přejde program k oknu s volbou typu výpočtu o kterém pojednává kapitola 2.5. Při volbě *No* budou použita stejná vstupní data, která byla zvolena při předchozí práci s programem. Není tedy nutné po novém startu programu vybírat znovu stejná data, se kterými se pracovalo naposledy.

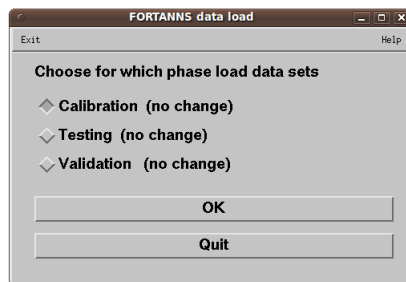


Obrázek 2: Dotaz na načítání dat

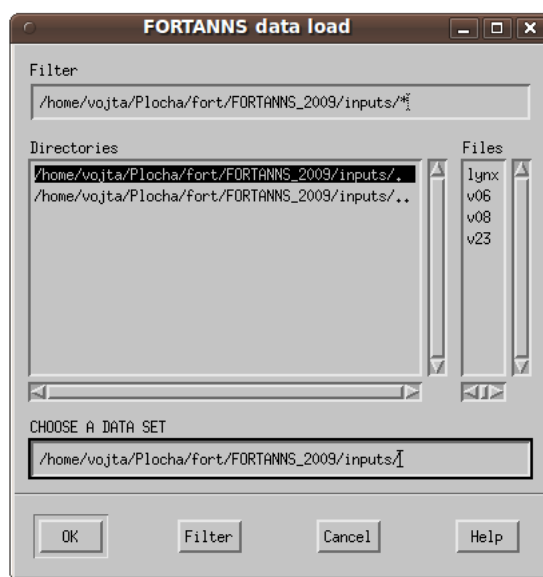
V případě volby *Yes* přejde program na okno s volbou fáze výpočtu, pro kterou budou datové soubory načítány, viz obr. 3.

Jednotlivé volby fází pro které budou načtena data jsou *Calibration*, *Testing* a *Validation*. V závorce za každou možností je zobrazen stav - *changed!*/*no change* - informující zda byly vybrány jiné vstupní datové soubory nebo zda zůstaly vstupní soubory pro danou fázi beze změn. Fáze výpočtu a načtení vstupních dat pro tyto fáze je úzce svázáno s možností volby fází při samotném nastavení výpočtu - pro fázi bez načtených data není možné spustit výpočet.

Vybrání fáze a potvrzení výběru kliknutím na *OK* posune program do okna, které slouží k volbě datových souborů (nahraných nejlépe ve složce **inputs**). Toto okno ukazuje obrázek 4.



Obrázek 3: Volba fáze pro kterou budou načtena data



Obrázek 4: Výběr datových souborů

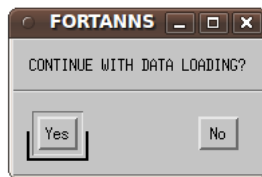
V tomto okně je možné volit soubory uložené v *inputs*. Soubory jsou vidět v pravém podokně s hlavičkou "Files". Ve složce *inputs* jsou nahrány zkušební časové řady. Soubor "lynx" obsahuje údaje o množství odchycených rysů v Kanadě v letech 1821 – 1934 (*Brockwell, Davis 1991*). Soubory "v06", "v08", "v23" jsou časové řady průtoků v hodinovém kroku.

Změna adresáře s vstupními soubory je možná v levém podokně s hlavičkou "Directories", ale pro pohodlnější práci je vhodnější pracovat se soubory nahranými do složky *inputs*. Horní podokno s hlavičkou "Filter" slouží k vyfiltrování požadovaných souborů podle názvu, filtrování se aktivuje kliknutím na *Filter*

Výběr souboru je potřeba potvrdit tlačítkem OK, kterým se aktivuje dialogové okno (obr. 5) s dotazem, zda pokračovat s výběrem dat pro aktuálně zvolenou fázi.

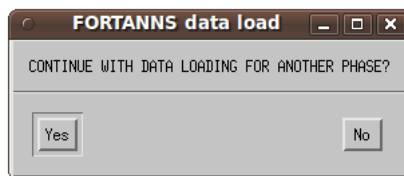
V okně pro výběr datových souborů (obr. 4) je možné vybrat vždy jeden soubor, ale každá fáze může mít víc vstupních datových souborů, proto v takovém případě je možné vybrat v aktuálním dialogovém okně (obr. 5) *Yes* a pokračovat ve výběru dalšího datového souboru pro danou fázi. Takto je možno zvolit libovolný počet vstupních souborů pro každou fázi.

Ukončení volby souborů pro danou fázi je možné kliknutím na *NO* v dialogovém okně pro pokračování výběru dat (obr. 5). Tímto se aktivuje další dialogové okno (obr. 6), ve kterém je možné zvolit pokračování načítání dat pro jinou fázi (*Yes*) nebo ukončit práci na výběru datových



Obrázek 5: Výběr datových souborů

souborů (*No*).

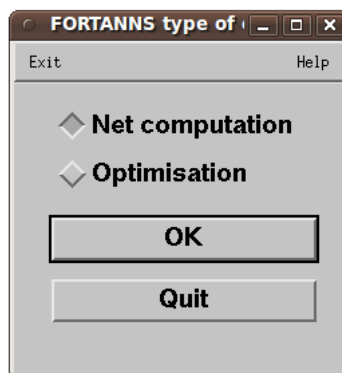


Obrázek 6: Výběr datových souborů pro jinou fázi

V případě výběru *Yes* se opakuje postup popsany výše pro další fáze, při volbě *No* přejde program k volbě typu výpočtu.

2.5 Volba typu výpočtu

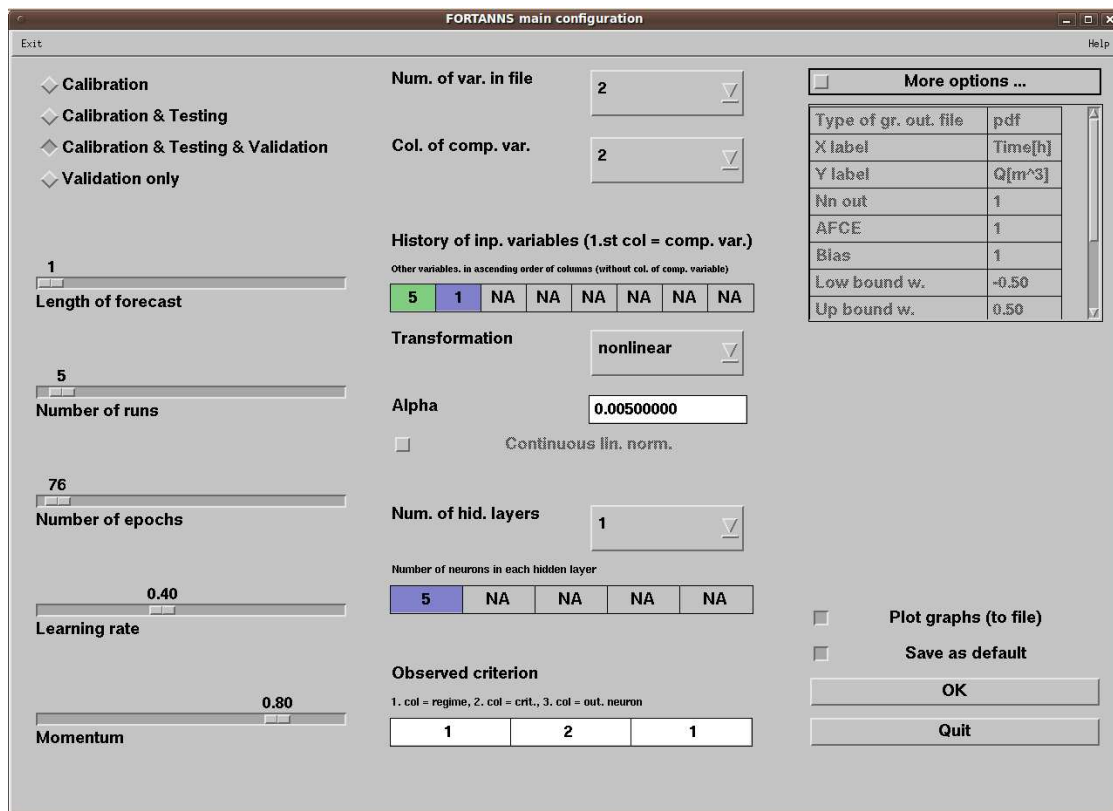
Obrázek číslo 7 ukazuje okno pro volbu výpočtu. Možnosti jsou *Net computation* - normální výpočet nebo *Optimisation* - mód optimalizace parametrů sítě. Výběr je třeba potvrdit pomocí *OK*. O jednotlivých volbách pojednávají následující kapitoly.



Obrázek 7: Volba typu výpočtu

3 Normální výpočet

Při volbě módu *Net computation* přejde program do hlavního okna pro nastavení architektury sítě, výpočtu, a výstupů. Toto hlavní okno ukazuje obrázek 8



Obrázek 8: Hlavní okno nastavení pro normální výpočet

Pro zjednodušení popisu jsou ovládací prvky logicky rozděleny do oblastí jak ukazuje obrázek č. 9. Další popis bude tedy postupně věnován ovládacím prvkům v oblastech 1 - 6.

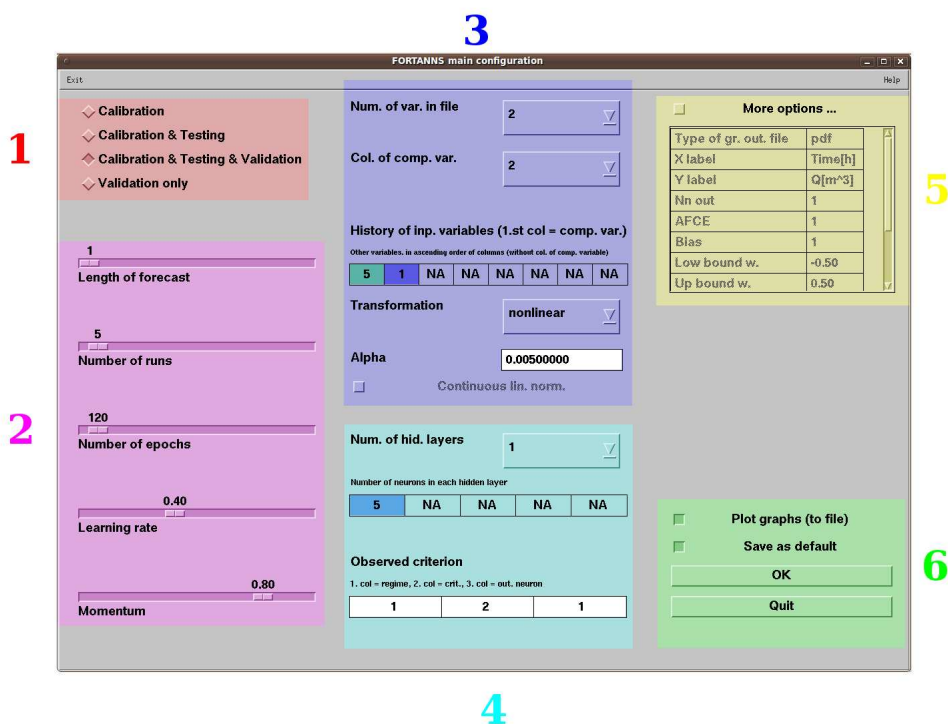
3.1 Nastavení fáze výpočtu (1. oblast)

Volba fáze určuje, zda program bude pracovat pouze s kalibračními daty nebo i s daty testovacími a validačními. Možnosti jsou *Calibration*, *Calibration & Testing*, *Calibration & Testing & Validation* a poslední možností je samostatná validace - *Validation*.

Nejdůležitější fází je kalibrace. Při kalibraci se neuronová síť učí (backpropagation learning algorithm) na datových vzorech závislosti a vztahy, které se v datech nachází. Tato fáze je časově nejnáročnější, po dokončení je k dispozici naučená neuronová síť (viz. kapitola 6), která může být použita v dalších fázích, nebo i k jiným účelům.

Testovací a validační fáze by měly mít pro výpočet rozdílná data, než jsou v každé jiné fázi. Testovací fáze (je-li zvolena) běží souběžně s kalibrační fází a analýzou výsledků obou fází se předchází takzvanému přeučení neuronové sítě. Tato kontrola přeučení je v současnosti ve vývoji. Stávající verze má testovací fázi fungující stejně jako validace - naučená síť je použita pro výpočet na datech, která nesloužila ke kalibraci.

Volba samotné validace slouží pro práci s již naučenou sítí, například pro použití sítě importované (ručně) z jiného programu nebo pro použití sítě na nových datech, která nebyla použita v žádné fázi při normálním výpočtu.



Obrázek 9: Rozdělení ovládacích prvků do oblastí

3.2 Volba parametrů sítě (2. oblast)

Mezi základní parametry modelu neuronových sítí implementovaného v programu FORTANNS patří nastavení délky předpovědi, počtu běhů programu, počtu epoch, hodnoty learning rate a hodnoty momenta. Volba hodnoty je prováděna tažením posuvníku, jemnější posun je možný směrovými klávesami.

Volba délky předpovědi určuje na jak dlouhý časový interval se program pokusí nalézt závislost mezi hodnotami vstupními (t_0, \dots, t_{-x} , kde x je historie dat, t_0 je hodnota v "přítomnosti") a hodnotou v budoucnosti.

Volba počtu běhů programu slouží k hrubému určení nejistot v předpovědi. Hodnota tohoto parametru určuje, kolik výpočtů sítě všech zvolených fází proběhne po startu programu. Při každém novém výpočtu je síť kalibrována s jinými inicializačními váhami v síti, takže při dostatečně vysokém počtu běhů programu je možné zhruba podchytit rozsah hodnot předpovědí modelu pro dané parametry sítě.

Počet epoch určuje, kolikrát dokola projde síť vzory vytvořené z dat. Jinými slovy určuje tento parametr, kolikrát při učení uvidí program konkrétní vzor. Hodnota parametru počtu epoch silně ovlivňuje kvalitu naučení sítě. Počet epoch je závislý na typu dat a ostatních parametrech, proto

je lepší pro jeho stanovení využít automatickou optimalizaci, viz. dále.

Hodnota momenta a learning rate určují rychlost a přesnost učení sítě. Tyto hodnoty jsou závislé na typu vztahů uvnitř dat.

Nalezení optimálních parametrů sítě je možné buď metodou pokus-omyl nebo, což je efektivnější, poskytuje program FORTANNS možnost optimalizace pomocí evolučních algoritmů (viz. kapitola 4). Tato optimalizace je zaměřena na počet epoch, momentum, learning rate a hodnotu alfa při nelineární transformaci (viz. kap. 3.3)

3.3 Volba parametrů vstupů (3. oblast)

První dva parametry v této oblasti určují počet sloupců v souboru se vstupními daty *Num. of var. in file* a pořadí sloupce s modelovanou proměnnou v tomto souboru *Col. of comp. var.*. Volba požadované hodnoty se uskuteční výběrem z rozbalovací rolety.

Počet sloupců ovlivňuje vzhled tabulky "History of inp. variables", která je v této oblasti jako další nastavení. Do této tabulky se volí historie vstupů, podle které bude počítána předpověď. První sloupec odpovídá vždy modelované proměnné, takže hodnota v tomto sloupci určuje počet vstupních hodnot té proměnné, která je modelována. Historie vysvětlujících proměnných je pak do tabulky zadávána v pořadí sloupců, jak jsou v souboru vstupních dat. Součet hodnot v této tabulce odpovídá počtu neuronů ve vstupní vrstvě neuronové sítě.

Příklad: Vstupní soubor se 3 sloupci proměnných - P,Q,T. Pro výpočet je požadována vstupní historie P = 1, Q = 2, T = 3. Modelovat je potřeba Q. V *Num. of var. in file* se tedy volí 3 (3 sloupce celkově ve vstup. souboru) a v *Col. of comp. var.* 2 (Q je 2. sloupec). V tabulce "History of inp. variables" pak 2, 1, 3 postupně zleva doprava.

Dalším nastavením je typ transformace vstupních dat. Neuronová síť vyžaduje hodnoty transformované na interval (0,1) nebo až do (-1,1) V rozbalovací roletě *Transformation* je možnost volby lineární nebo nelineární transformace. V závislosti na volbě typu transformace se aktivuje nastavení alfy - nelineární typ, nebo nastavení zapnutí/vypnutí postupné lineární transformace. Definici nelineární transformace uvádí rovnice č. 1. Lineární transformace je realizována podle vztahu 2. Rozsah lineární transformace je možné upravovat, viz. kapitola 3.5. Výsledné hodnoty jsou po ukončení výpočtu retransformovány.

$$x_t^T = 1 - e^{-alfa * x_t} \quad (1)$$

$$x_t^T = l_{min} + \frac{(l_{max} - l_{min})(x_t - x_{min})}{x_{max} - x_{min}} \quad (2)$$

kde	x_t^T	je transformovaná hodnota proměnné
	$alfa$	je volitelný parametr nelineární transformace
	x_t	je pozorovaná hodnota proměnné
	l_{min}	je spodní hranice rozsahu hodnot pro lin. transformaci
	l_{max}	je horní hranice rozsahu hodnot pro lin. transformaci
	x_{min}	je minimální hodnota proměnné v souboru vstupních dat
	x_{max}	je maximální hodnota proměnné v souboru vstupních dat

3.4 Volba vnitřní architektury sítě a sledování výstupů (4. oblast)

V této oblasti je jako prvním nastavením počet skrytých vrstev neuronové sítě (*Num. of hidd. layers*). Volba hodnoty se provede z rozbalovací rolety a tato hodnota odpovídá počtu vrstev mezi vstupní a výstupní vrstvou. Pro reálné využití není potřeba volit více než 2 vrstvy, pro většinu aplikací však stačí jen 1 vrstva.

Volba počtu vrstev ovlivní i následující tabulku ("Number of neurons in each layer"), kde hodnoty v jednotlivých sloupcích odpovídají počtu neuronů v dané vrstvě.

Posledním objektem v této oblasti je tabulka "Observed criterion" se 3 sloupci pro nastavení sledování kvality výstupů. První sloupec v této tabulce je dotaz na fázi, pro kterou bude kritérium sledováno. Vpisovat je možné čísla 1, 2, 3, tyto čísla odpovídají postupně fázím kalibrace, testování, validace. Volba sledované fáze musí být uskutečnitelná v závislosti na volbě fáze výpočtu (kap. 3.1). Volba sledování kritéria ovlivňuje výpis na obrazovku při výpočtu, výpis modelovaných dat a residuí do souboru a grafické výstupy (kapitola 6).

Druhý sloupec tabulky odpovídá typu kritéria, které bude sledované. Možnosti nastavení jsou dvě - číslo 1 odpovídá Nash-Sutcliffeho koeficientu, číslo 2 odpovídá indexu persistence. Nash-Sutcliffeův koeficient (koeficient determinace) je definován v rovnici 3. Index persistence je definován v rovnici 4

$$NASH = 1 - \frac{\sum_{t=1}^T (x_t - \hat{x}_t)^2}{\sum_{t=1}^T (x_t - \bar{x})^2} \quad (3)$$

$$PERI = 1 - \frac{\sum_{t=1}^T (x_t - \hat{x}_t)^2}{\sum_{t=1}^T (x_t - x_{t-p})^2} \quad (4)$$

kde x_t je pozorovaná hodnota proměnné
 \hat{x}_t je predikovaná hodnota proměnné
 \bar{x} je střední hodnota proměnné
 x_{t-p} je pozorovaná hodnota proměnné v čase $t - p$, kde p je délka předpovědi

Z definicí kritérií je patrné, že čím blíže leží hodnota kritérií k jedné, tím je daný model lepší. NASH kritérium je míra o kolik lepší je daný model, než kdyby byla predikovaná hodnota modelována pomocí střední hodnoty proměnné a PERI kritérium ukazuje o kolik je lepší daný model, než kdyby byla predikovaná hodnota modelována pomocí poslední známé hodnoty vstupů.

Poslední sloupec tabulky "Observed criterion" odpovídá počtu výstupních neuronů - je-li volena jiná hodnota než jedna, tak bude program počítat víceokrou predikci (která však musí být nastavena také v pokročilém nastavení, viz. kap. 3.5). Program však efektivněji pracuje v módu jednokrokové predikce.

3.5 Pokročilé nastavení (5. oblast)

Pátá oblast zahrnuje pokročilá nastavení programu. Tabulka pro změnu těchto nastavení se aktivuje zapnutím tlačítka *More options...*

První tři řádky tabulky odpovídají nastavení pro grafické výstupy. Grafické výstupy je možno uvažovat jen v případě, že je nainstalován program R (<http://www.r-project.org/>). V prvním řádku tabulky je možné měnit typ grafického souboru, možné typy jsou *pdf* a *postscript*. Druhý a třetí řádek slouží pro změnu popisků os v grafech (max délka řetězce 20 znaků bez mezer).

Řádek "Nn out" slouží pro volbu počtu neuronů ve výstupní vrstvě. Jinou hodnotou než jedna je nastavena víceokrou predikce. To znamená, že program bude počítat s modelem neuronové sítě, která predikuje najednou x_{t+1} , x_{t+2} , ..., x_{t+ms} , kde ms je počet výstupních neuronů. Tento typ predikce však není tak efektivní jako v případě jednokrokové predikce.

Řádek AFCE slouží pro volbu aktivační funkce neuronu. Možnosti jsou čtyři. Číslo 1 odpovídá hyperbolické tangentě (rovnice 5), číslo 2 je sigmoida (rovnice 6), 3 odpovídá Gaussově funkci (rovnice 7) a 4 patří funkci arkustangens (rovnice 8)

$$AFCE = \tanh(i) = \frac{e^{2i} - 1}{e^{2i} + 1} \quad (5)$$

$$AFCE = \frac{1}{1 + e^{-i(slope)}} \quad (6)$$

$$AFCE = e(-i^2) \quad (7)$$

$$AFCE = \frac{2}{\pi} \arctan(i) \quad (8)$$

kde	i	je vstup do neuronu
	e	je základ přirozeného logaritmu (Eulerovo číslo)
	$slope$	je nastavitelný parametr sklonu sigmoidy
	π	je Ludolfovo číslo

Řádek "Bias" kontroluje zapnutí (číslo 1) nebo vypnutí (číslo 0) biasu ve výpočtu neuronové sítě.

Řádek "Low bound w." a "Up bound w." slouží pro nastavení rozmezí hodnot vah, které jsou na začátku výpočtu kalibrační fáze nastaveny náhodně v daném rozmezí.

Řádky "l_min" a "l_max" jsou určeny pro volbu rozmezí transformovaných hodnot při lineární transformaci (viz. kap.3.3 rovnice 2).

V předposledním řádku tabulky pokročilého nastavení je volba "Const. w." Jedná se o volbu zapnutí (číslo 1) nebo vypnutí (číslo 0) konstantních vah, neboli zapnutí/vypnutí náhodného nastavení vah na začátku výpočtu. Je-li zvolena 0 jsou váhy náhodně generovány, pro volbu 1 je potřeba mít soubor s váhami připravený v souboru **config/vahy_k** a tento soubor pak bude sloužit k výpočtu.

Poslední řádek "Slope" slouží k nastavení parametru sklonu sigmoidální funkce. Toto nastavení má efekt pouze v případě volby "AFCE" = 2.

3.6 Potvrzení nastavení (6. oblast)

V této oblasti se kromě tlačítka *OK* pro potvrzení nastavení a tlačítka *QUIT* pro ukončení programu nachází dvě další volby. Volba *Plot graphs (to file)* je určena pro zapnutí/vypnutí vykreslování grafů do souboru. Zapnutí této volby má efekt pouze při nainstalovaném programu R. Dále je třeba upozornit, že při volbě většího počtu běhů programu (kap. 3.2) zabere generování grafiky dost času, takže je potřeba dostatečná trpělivost nebo je třeba tuto volbu vypnout. Druhou volbou je možnost uložení nastavení *Save as default*. Zapnutí této volby zaručí, že po potvrzení tlačítkem *OK*, bude současné nastavení uloženo pro další start programu nebo volbu nastavení.

4 Optimalizace

Zvolením typu výpočtu *Optimisation* v okně volby výpočtu (kap. 2.5) se nashoduje okno s nastavením velice podobným vzhledem jako při normálním výpočtu (obr. 10). V oblasti 1 (obr. 9) ubyla možnost volby samostatné validace. Oblasti 3, 4 jsou bez výraznějších změn, parametr alfa je optimalizován, takže jeho volby v oblasti 3 nemá praktický význam. Velmi důležitá je při optimalizaci tabulka "Observed criterion" (obl. 4) pro sledování vstupů. Zatímco v předchozím typu výpočtu byla opravdu jen pro sledování kvality výpočtu, v optimalizaci je její nastavení důležité, protože nastavená hodnota k sledování bude zároveň hodnotou pro výpočet fitness každého řešení a tím zároveň bude tato hodnota rozhodovat o síle daného řešení v selekci pro křížení. V oblasti 5 byly smazány možnosti, které nemají při optimalizaci efekt.

Největších změn však dostala oblast 2. V této oblasti je kromě nastavení délky předpovědi vše jiné. Při optimalizaci se na tomto místě nachází hlavní nastavení evolučního výpočtu. Jedná se o dvě posuvné lišty a čtyři tabulky.

První posuvník nastavuje počet jedinců (řešení) kteří budou generováni ve startovní generaci s náhodně nastavenými optimalizovanými parametry. Druhým posuvníkem se kontroluje nastavení počtu generací, tj. kolikrát dojde k výpočtu fitness, křížení jedinců a vytvoření potomků.

Prvním tabulkovým nastavením je rozsah epoch ve kterém bude náhodně volena hodnota počtu epoch pro každého jedince ve startovní generaci. Zadává se spodní hranice, horní hranice a jednotky na které bude náhodně generovaná hodnota zaokrouhlena.



Obrázek 10: Okno pro nastavení optimalizace

Druhá tabulka slouží k nastavení rozsahu hodnot ve kterých bude program optimalizovat parametry alfa, learning rate a momentum.

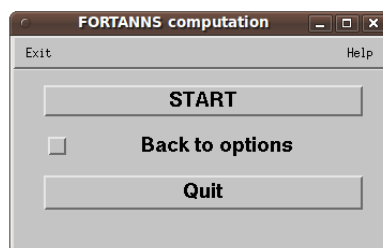
Tabulka uvedená "Ratio of survival" udává poměr, jaká část nejlepších jedinců v každé generaci projde procesem křížení.

Poslední Tabulka "Number of ind. showed in file" určuje kolik nejlepších jedinců se svými optimalizovanými parametry bude vypsáno do souboru.

Zvolené nastavení se potvrdí tlačítkem *OK* vpravo dole v okně pro nastavení optimalizace.

5 Start výpočtu

Poslední okno určené pro start výpočtu je uvedena na obrázku 11. V tomto stručném okně je možno spustit výpočet (*Start*), vrátit se do volby nastavení (*Back to option*) nebo program ukončit. Okno pro start výpočtu zůstane otevřené i po ukončení výpočtu, takže je možné se v případě potřeby rychle vrátit do nastavení výpočtu a nastartovat program s jiným nastavením.



Obrázek 11: Okno pro start výpočtu

6 Výstupy

Pro výstupy vytváří program FORTANNS při každém spuštění složky *optim*, *kal*, *tes*, *val*. Staré výstupy jsou smazány, čili je vhodné před dalším spuštěním programu kopírovat již vypočítané výstupy do jiného adresáře. Uvedené hlavní složky výstupů obsahují další adresáře a složky, které jsou uvedeny v následujících podkapitolách.

6.1 Výstupy kalibrace

Výstupy kalibrace jsou uloženy ve složce *kal*. Obsahem tohoto adresáře je složka *grafy* a složky *m1*, *m3*, ..., *m_x*, kde *x* je počet běhů programu.

Výstupní soubory v hlavní složce *kal* jsou:

data_best_kal, kde jsou uložena měřená a modelovaná data pro kalibraci poskytnutá nejlepším modelem

kriteria_kal, kde jsou uloženy hodnoty kritérií pro každý model (běh programu)

residuals_best_kal, kde jsou residua pro predikci nejlepším modelem

stat_krit_kal, kde jsou maximální, minimální a průměrné hodnoty kritérií

vahy_best, kde jsou uloženy nakalibrované váhy modelu neuronových sítí

vzory_kal, kde jsou uloženy jednotlivé vzory pro neuronovou síť vytvořené z dat

Výstupní soubory ve podadresáři *kal/grafy* jsou:

ACFres_kal_n1.pdf - graf autokorelační funkce

PACFres_kal_n1.pdf - graf parciální autokorelační funkce

NS_crit_kal_n1.pdf - průběh hodnot Nashova kritéria pro jednotlivé modely

PI_crit_kal_n1.pdf - průběh hodnot indexu persistence pro jednotlivé modely

scatt_bestkal_n1.pdf - bodový graf modelované vs. měřené hodnoty

SimObs_kal_n1.pdf - graf průběhu měřených a modelovaných hodnot časové řady

scr_kal - R skript pro grafy

V jednotlivých podadresářích *m_x* je soubor **residuals** s residuii pro konkrétní model, soubor **vahy** s nakalibrovanými váhami pro konkrétní model (běh programu) a soubor **data**, kde jsou uložena měřená a modelovaná data pro konkrétní model.

6.2 Výstupy testování a validace

Výstupy kalibrace jsou uloženy ve složce *kal*. Typy výstupů těchto fází jsou totožné s fází kalibrace, jen soubory s váhami zde chybí, jelikož váhy jsou vstaženy pouze ke kalibraci.

6.3 Výstupy optimalizace

Výstupy kalibrace jsou uloženy ve složce *optim*. Výstupem optimalizace jsou čtyři soubory.

Soubor **best_opt** obsahuje optimalizované parametry pro zvolený počet nejlepších jedinců

Soubor **evolution.pdf** je grafický výstup ukazující vývoj sledovaného kritéria (max., min., prům.,) v jednotlivých generacích

Soubor **scr_opt** je R skript pro grafy

Soubor **stat_opt** obsahuje maximální, minimální a průměrnou hodnotu sledovaného kritéria v každé generaci.

7 Příklady

Následující dvě podkapitoly se věnují ukázkovému použití programu FORTANNS. Vstupní data pro výpočet jsou zkušební časové řady, přednahrané ve složce *inputs*

První příklad se zabývá modelováním předpovědí z průtokových dat. Vstupními daty jsou tři povodňové vlny, které mají rozdílnou velikost. Tyto vlny jsou ze stejného povodí. Počítána bude kalibrace, která bude opravována ručně pomocí testovacích dat a vybraný model bude nakonec vyzkoušen na validačních datech.

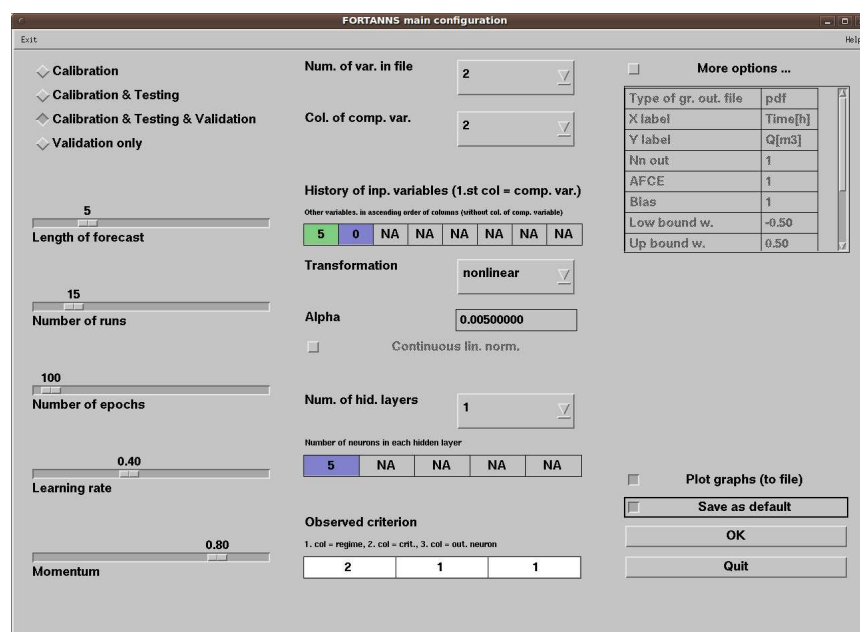
Druhý příklad je věnován predikci počtu jedinců rysa kanadského (*Lynx canadensis*). Vstupními daty je stoletá řada počtu jedinců chycených v Kanadě v letech 1821 – 1934. V tomto příkladě bude počítána pouze kalibrace spolu s předchozí optimalizací parametrů sítě. Tato data nejsou typickým typem dat pro modelování neuronovými sítěmi, avšak jako příklad je možné tato data použít. Z výsledků dále je vidět, že program i pro tato data bez předchozích úprav poskytuje relativně uspokojivou předpověď.

Uvedené příklady jsou jen ilustrační a pro demonstraci ideálního fungování programu nedostatečné, zejména z důvodu malé délky vstupních časových řad.

7.1 Predikce odtoku

Načtení dat: Pro predikci odtoku bude sloužit soubor **v06** jako kalibrační, soubor **v08** jako testovací a **v23** jako validační. Data byla načtena podle pokynů v kapitole 2.4.

Nastavení výpočtu: Typ výpočtu je volen *Net computation*. Další nastavení je možné podle vlastního uvážení, pro lepší výsledky bude nutno vyzkoušet víc variant nastavení. Pro účely tohoto příkladu bylo navoleno nastavení, které ukazuje obr. 12.



Obrázek 12: Nastavení programu pro přepověď průtoků

Výsledky: Po startu programu jsou vypisovány do okna průběžné výsledky, ale hlavní výstupy jsou vepsány do souborů. Ve složkách *kal*, *tes*, *val*, které se vytvořily v adresáři FORTANNS jsou uloženy informace o průběhu a kvalitě výpočtu. Strukturu těchto složek popisuje kap. 6. Pro účely tohoto příkladu jsou vybrány jen nejdůležitější výstupy, kterými jsou hydrogramy

měřených a modelovaných dat (obr. 15 - obr. 17) a průběh hodnot kritérií pro jednotlivé modely (obr. 18 - obr. 20). Zobrazené výstupy jsou podřízeny nastavenému výběru nejlepšího modelu z testovací fáze, takže podle obr. 19 a vypočtených kritérií je nejlepší model č. 6. Použití vah nakalibrovaným pro tento model v kalibrační fázi odpovídá i červená modelovaná čára v hydrogramech (obr. 15 - obr. 17) pro každou fázi.

V následující tabulce jsou uvedené maximální, minimální a průměrné hodnoty obou kritérií pro všechny tři fáze.

Fáze	Kalibrace			Testování			Validace		
Hodnota	max.	min.	prům.	max.	min.	prům.	max.	min.	prům.
NASH	0.988	0.874	0.971	0.972	0.897	0.960	0.986	0.933	0.977
PERI	0.612	-3.011	0.064	0.519	-0.803	0.300	-0.125	-4.555	-0.906

Z uvedených výsledků vyplývá, že pro kalibrační i testovací fázi měl vypočítaný model dobré výsledky, maximální hodnoty kritérií byly v obou fázích vyšší než 0.5, což je pro pětihodinovou predikci průtoků velmi slušné. Takto nakalibrovaný model však již nepotvrdil své kvality ve fázi validační (přestože parametr NASH je na velmi vysoké hodnotě). Index persistence ukazuje, že pro modelování odtoku na tomto datovém souboru by bylo v průměru výhodnější použít poslední známou hodnotu vstupu a tuto hodnotu brát jako modelovaný výstup. Z takto získaných výsledků je zřejmé, že by bylo vhodné upravit nastavení modelu a změnit například typ transformace nebo počet epoch aj., aby došlo k vylepšení předpovědi.

7.2 Predikce velikosti populace ryba kanadského

Načtení dat: Pro predikci odtoku bude sloužit soubor **lynx**. V tomto příkladě bude počítána pouze kalibrace, takže se celý soubor načte do datových souborů pro kalibrační fázi. Data byla načtena podle pokynů v kapitole 2.4.

Nastavení výpočtu 1 - optimalizace: Jako typ výpočtu je nejprve voleno *Opimisation*. Další nastavení je možné podle vlastního uvážení, čím větší rozsah parametrů (zejména epoch a počtu jedinců) tím bude trvat výpočet optimalizace déle. Pro účely tohoto příkladu bylo navoleno nastavení, které ukazuje obr. 13.

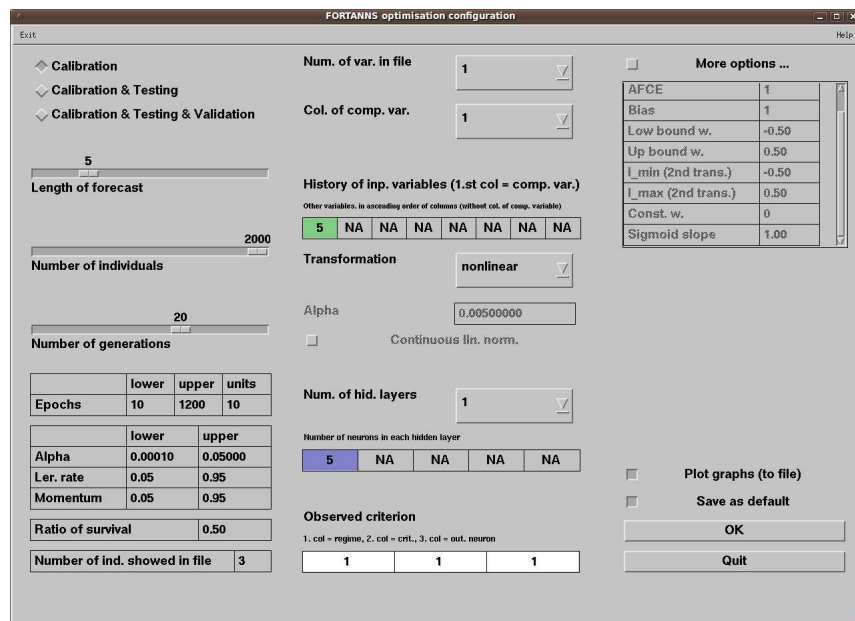
Výsledky 1 - optimalizace: Po provedení výpočtu (1h 35min) byla vytvořena složka **optim**, ve které jsou soubory s optimalizovanými parametry a graf vývoje řešení (obr. 21)

Parametry které odpovídají 3 nejlepším řešením jsou v následující tabulce.

Řešení	epochy	alfa	learning rate	momentum	NASH
1.	1190	$4.569 \cdot 10^{-4}$	0.057	0.744	0.653
2.	1140	$4.569 \cdot 10^{-4}$	0.079	0.670	0.647
3.	1030	$4.406 \cdot 10^{-4}$	0.118	0.472	0.637

Z tabulky je možné odečíst přesné hodnoty parametrů, ale pro vlastní výpočet bude voleno nastavení optimalizovaných parametrů zhruba následující: 1100 epoch, $4.57 \cdot 10^{-4}$ parametr transformace alfa, 0.08 learning rate a 0.7 momentum (zaokrouhlené přibližné průměrné hodnoty nebo dle citu volené).

Nastavení výpočtu 2 - výpočet predikce: Po novém spuštění programu je typ výpočtu volen *Net computation*. Další nastavení odpovídá nastavení při optimalizaci a hodnotám optimalizovaných parametrů (obr. 14), jak bylo uvedeno výše.



Obrázek 13: Nastavení programu pro optimalizaci

Výsledky 2 - výpočet predikce: Průběh měřených a predikovaných hodnot pro pětiletou predikci ukazuje obr. 22.

V následující tabulce jsou uvedené maximální, minimální a průměrné hodnoty obou kritérií pro fázi kalibrace.

Fáze	Kalibrace		
Hodnota	max.	min.	prům.
NASH	0.540	0.300	0.435
PERI	0.858	0.784	0.825

V hodnotách kritérií na rozdíl od prvního příkladu byly zaznamenány nižší hodnoty v Nashově kritériu a index persistence je naopak vyšší. Pozitivní skutečnost je, že ani jedno kritérium nekleslo pod nulu, z čehož se dá usuzovat na dobře optimalizované parametry. Z obrázku 22 je patrné, že model poskytoval dobré výsledky v přibližně středních stavech, podchytit nejvyšší stavy se modelu nepodařilo.

FORTANNNS main configuration

Exit Help

- Calibration
- Calibration & Testing
- Calibration & Testing & Validation
- Validation only

Length of forecast: 5

Number of runs: 20

Number of epochs: 1100

Learning rate: 0.08

Momentum: 0.70

Num. of var. in file: 1

Col. of comp. var.: 1

History of inp. variables (1.st col = comp. var.)

Other variables, in ascending order of columns (without col. of comp. variable)

5	NA	NA	NA	NA	NA	NA
---	----	----	----	----	----	----

Transformation: nonlinear

Alpha: 0.000457

☐ Continuous lin. norm.

Num. of hid. layers: 1

Number of neurons in each hidden layer

5	NA	NA	NA	NA
---	----	----	----	----

Observed criterion

1. col = regime, 2. col = crt., 3. col = out. neuron

1	1	1
---	---	---

More options ...

Type of gr. out. file	pdf
X label	Time[ye
Y label	Number
Nn out	1
AFCE	1
Bias	1
Low bound w.	-0.50
Up bound w.	0.50

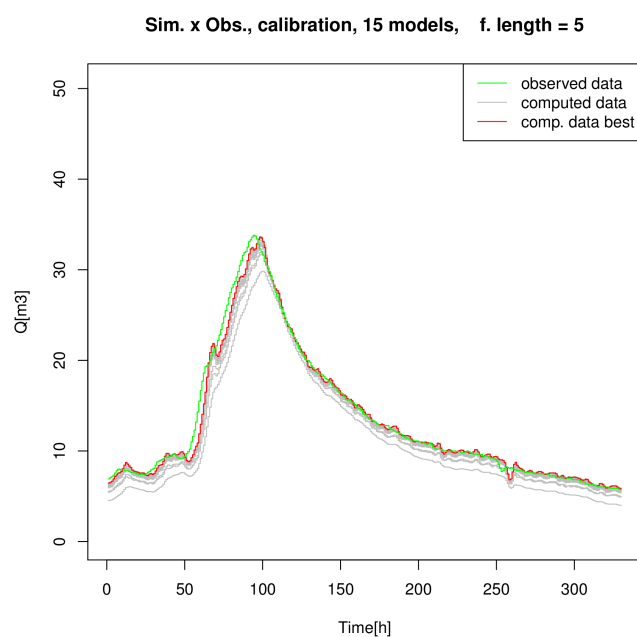
☐ Plot graphs (to file)

☐ Save as default

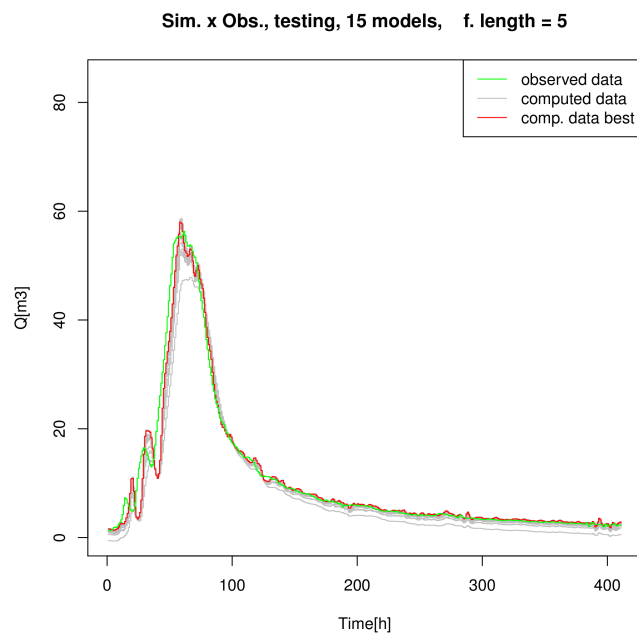
OK

Quit

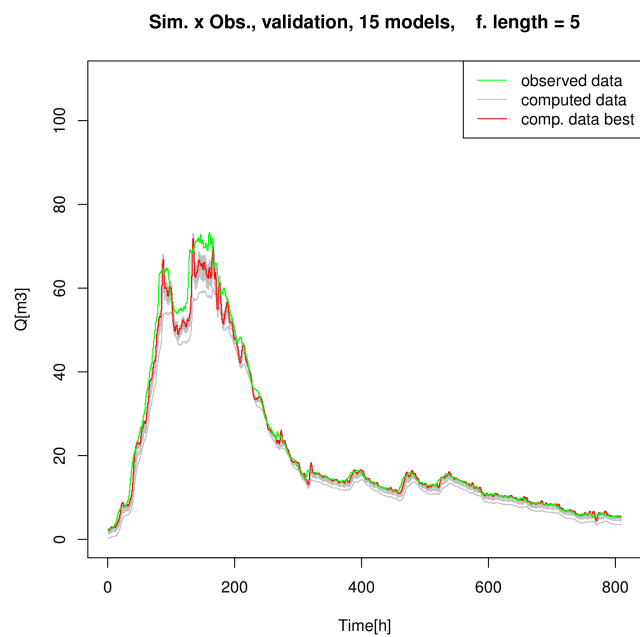
Obrázek 14: Nastavení programu pro výpočet



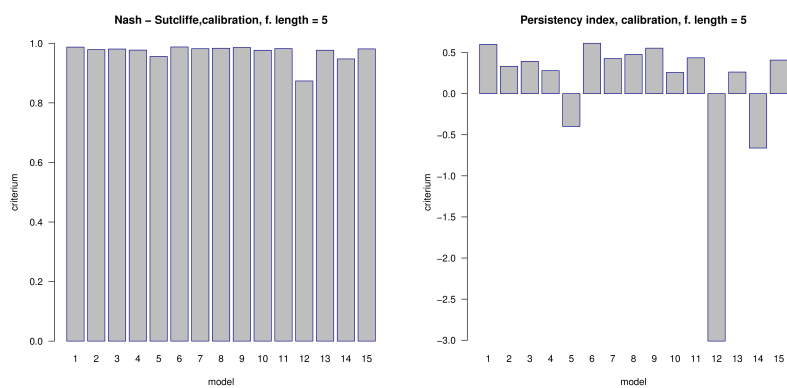
Obrázek 15: Měřené a modelované hodnoty průtoků. Kalibrace



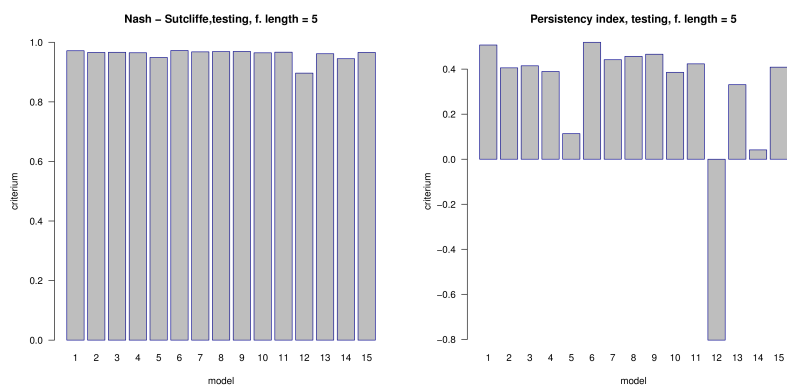
Obrázek 16: Měřené a modelované hodnoty průtoků. Testování



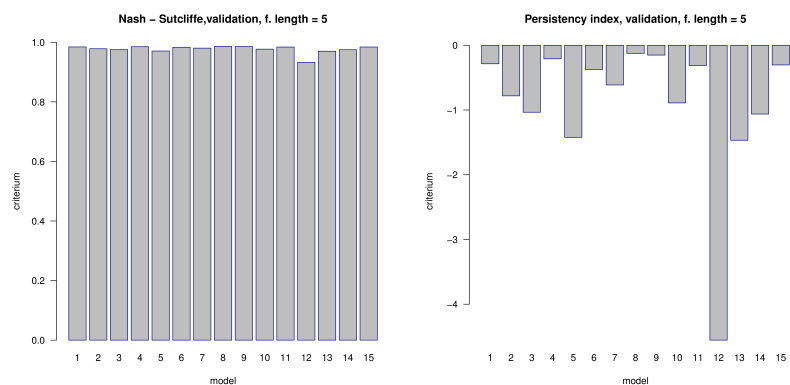
Obrázek 17: Měřené a modelované hodnoty průtoků. Validace



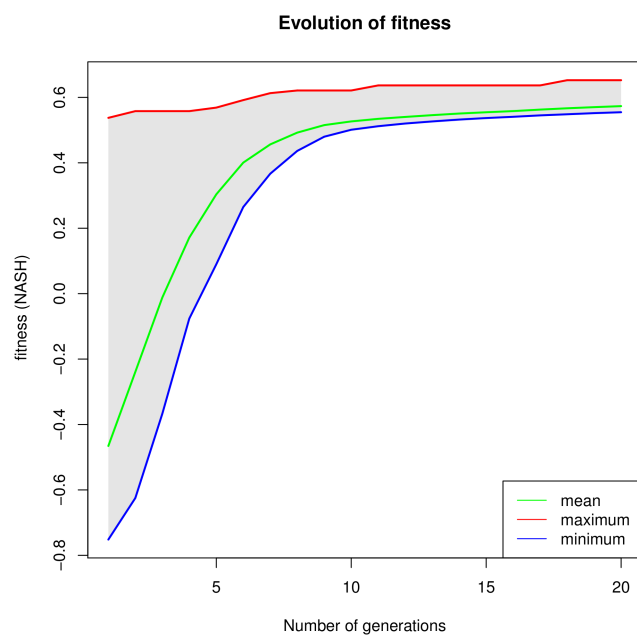
Obrázek 18: Nashovo kritérium a index persistence. Kalibrace



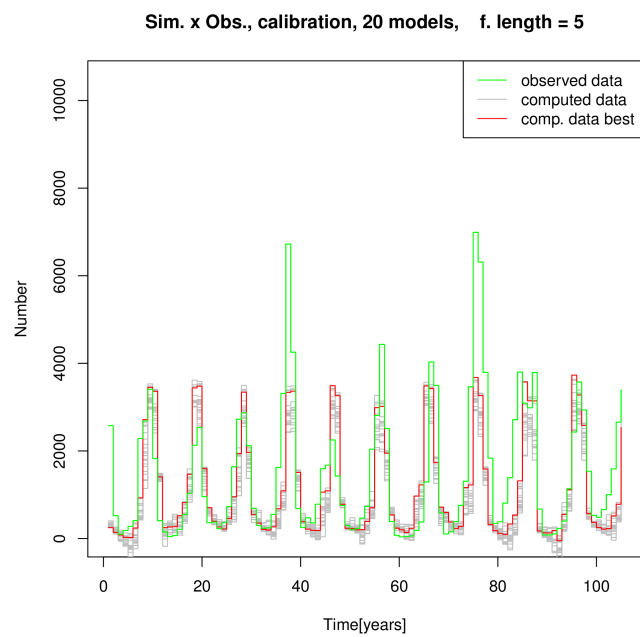
Obrázek 19: Nashovo kritérium a index persistence. Testování



Obrázek 20: Nashovo kritérium a index persistence. Validace



Obrázek 21: Graf vývoje jednotlivých generací řešení



Obrázek 22: Měřené a modelované počty jedinců. Kalibrace

Literatura

Brockwell P. J. and Davis R. A., 1991: *Time Series and Forecasting Methods*, Springer, série G (str. 557)

Fausett L. V., 1994: *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*, Pearson Education, 461s

Hynek J., 2008, *Genetické algoritmy a genetické programování*, Grada, 200s

Onwubolu G. C., Babu B. V., 2004: *New Optimization Techniques in Engineering*, Springer, 712s

Šíma J., Neruda R., 1996, *Teoretické otázky neuronových sítí*, MATFYZPRESS, Praha, 390s

Zelinka I., Oplatková Z., Šeda M., Ošmera P., Včelař F., 2008, *Evoluční výpočetní techniky - principy a aplikace*, BEN, 536s