

# ***L 12    Statistická adaptace***

***Oddělení numerické předpovědi počasí***

***© ČHMÚ 2007***

# Plán přednášky

- Úvod
- Přehled rozpracovaných metod (teorie)
- Aplikace metod v ČHMÚ
- Verifikace (ukázka pro hydrologické povodí)

# Statistická adaptace

- *patří k ní všechny metody, pomocí kterých se numerická předpověď modelu dostává v automatické formě k uživateli:*

- *filtrační metody*

- *inter-extrapolační metody*

- *regresní metody*

- *pravděpodobnostní metody*

- *jejich kombinace*

# Směřování

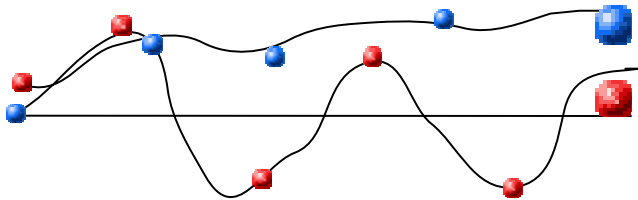
- přechod k multivariačním metodám
- umožňují pracovat s kratšími vstupními soubory, umí si lépe poradit se závislými prediktory
- zkracování vstupních souborů:
  - filtrační metody
  - přechod od dlouhé serie (5.2002 u nás) ke kratší přes snížení počtu prediktorů, přes využití multivariačních metod
- využití informace v prostoru:
  - prediktorem je jenom předpovídaný prvek
  - prediktory jsou předpovědi modelu ve více bodech nad měřením
  - snaha o využití informace nejen o velikosti předpovídaného prvku, ale i o jeho proměnlivosti

# Rozpracované metody (teorie)

- Kalmanův filtr - Maximální a Minimální teplota
- BMA (Bayesian Model Averaging) - srážky
- MOS (regrese) - rychlost větru
- PLSR (Partial Least Squares Regression) - rychlost větru

# Kalmanův filtr

- vypočte odchylku předpověď – měření a její rozptyl za 8 dní
- předpovídá další odchylku, větší váhu dává na poslední dny, zohledňuje její rozptyl (v případě většího rozptylu zmenšuje předpovídanou odchylku)
- řeší případy, kdy je počasí několik dní stejné a chyba modelu se opakuje
- řeší případy, kdy dochází k lokální chybě modelu, něco v modelu není zachyceno přesně (lokální půdní model, lokální orografie, lokální efekty počasí)
- pro dva případy z obrázku pro odchylku dobře pracuje v modrém případě, v červeném případě nedělá skoro nic



# BMA

- průměrovací metoda (nehledá nejlepší prediktory jako regrese, ale nejlepší kombinaci prediktorů)
- v principu dělá vážený průměr z prediktorů, optimální váhy pro jednotlivé prediktory hledá pomocí „Maximum Likelihood Metody“ (metoda maximální věrohodnosti)
- v prvním kroku dělá pravděpodobnostní předpověď pro zvolený práh, v druhém kroku podmíněnou deterministickou předpověď pro vybranou pravděpodobnostní třídu
- **výhody:**
  - zohledňuje náhodnost v předpovědi
  - dobře pracuje se závislými prediktory
  - dobře pracuje s nekalibrovanými prediktory
  - dobře pracuje s prediktory s různým rozdělením
  - tradičně se využívá na zpracování ansámblových předpovědí

# MOS

- lineární regresní model vytvořený pomocí metody postupné krokové regrese z jednotlivých prediktorů
- vytváří se na delším souboru modelových výstupů a pozorování (u nás serie předpovědí od května 2002 v současném modelovém rozlišení)
- postupná kroková regrese vybere prediktory, které nejlépe upravují předpověď modelu a vypočte jejich koeficienty
- regresní modely jsou aplikovány na aktuální modelové výstupy
- tradiční metoda, u nás v provozu 06/2000-03/2003 pro termínové teploty, maximální a minimální teploty, rychlost větru, oblačnost



# PLSR

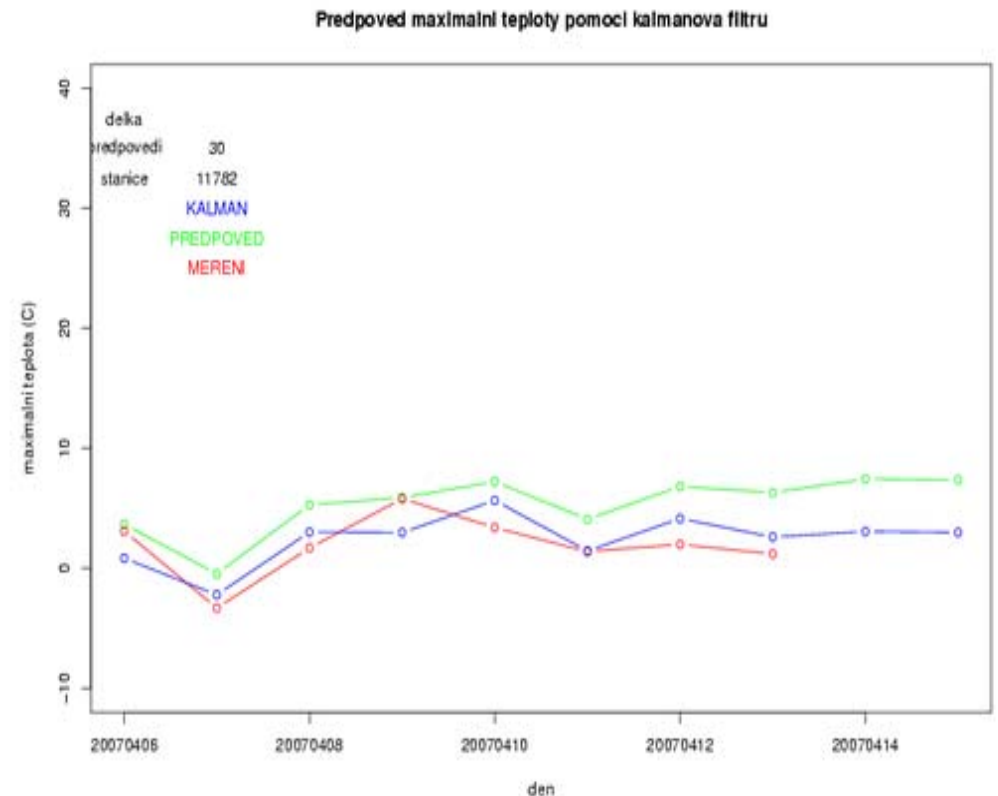
- postupuje podobně jako analýza hlavních složek (PCA), vytváří nové prediktory z původních proměnných, které jsou na sobě nezávislé
- z nových proměnných vytváří regresní model pomocí krokové vícerozměrné regrese
- v každém kroku postup opakuje, po odečtení prvního nového prediktoru hledá nový soubor navzájem nezávislých prediktorů, které by nejlépe vysvětlovaly kovarianci mezi prediktorem a prediktandem
- nezaručuje úplnou nezávislost prediktorů
- kombinuje výhody PCA a krokových regresí
- pracuje dobře se závislými prediktory
- pracuje dobře na kratších souborech předpovědí a měření

# Aplikace metod v ČHMÚ

- Předpověď maximální a minimální teploty pro synoptické stanice
- Předpověď srážkových úhrnů pro hydrologická povodí
- Předpověď rychlosti větru pro letiště
- ? (podle dalších požadavků uživatelů)

# Předpověď maximální a minimální teploty pro synoptické stanice

- metoda: Kalmanův filtr, pro 31 synoptických stanic
- zkušební poloprovoz na: <http://voodoo.chmi.cz/~mma105/prohlizec.html>
- z testovaných období dal Kalmanův filtr nejlepší výsledky u stabilní situace 11.-21.4. pro většinu stanic (systematická chyba se objevila na více stanicích)
- pro stanici na Lysé Hoře upravuje předpověď často (pravděpodobná systematická lokální chyba orografie nebo půdního modelu)



# Předpověď srážkových úhrnů pro hydrologická povodí (1)

- prvním řešeným problémem bylo vybrat vhodnou kombinaci bodů z modelu ALADIN nad zadaným hydrologickým povodím
- ukázalo se, že průměr z bodů nad povodím není nejvhodnější, testovaly se nad daným povodím různé výběry bodů
- pokusili jsme se využít metody statistické adaptace, aby si samy vybíraly body nad povodím (výhodou tohoto postupu bylo, že metody mohly vedle srážkového úhrnu využívat i informaci o proměnlivosti pole nad danou oblastí)
- pro pravděpodobnostní předpověď srážkových úhrnů jsme srovnali metody: Multivariační diskriminační analýza, Logistické regrese, BMA
- pro podmíněnou kvantitativní předpověď srážkových úhrnů jsme srovnali metody: Barisotropické regrese, PLSR, BMA

# Předpověď srážkových úhrnů pro hydrologická povodí (2)

- nejlepší výsledky dala metoda BMA ve svém nestandardním využití (standardně se využívá na zpracování ansámblových předpovědí)
- v našem případě: „pseudo-ansámbli“ z jednotlivých bodů modelu ALADIN nad daným hydrologickým povodím z jednoho běhu modelu ALADIN
- dává oba druhy předpovědi: pravděpodobnostní předpověď a podmíněnou kvantitativní předpověď srážkových úhrnů větších než zvolený práh
- jako jediná byla schopná pracovat s vyššími prahy srážkových úhrnů kolem 1, 2, 5, .... mm / 6hodin
- v současnosti probíhá srovnání málo se lišících modifikací BMA a výběr vhodných prahů pro vybraná hydrologická povodí

# Předpověď rychlosti větru pro letiště

- **pravděpodobný postup:**
  - MOS od května 2002 pro nejbližší bod k letišti ze všech prediktorů
  - PLSR od května 2002 jen pro rychlost větru v okolí  $4 \times 4 = 16$  bodů kolem letiště (závislé prediktory) = pokus vzít informaci jen z pole větru a jeho proměnlivosti kolem stanice
  - srovnání MOS a PLSR (pokud bude PLSR úspěšné, pokus zkrátit délku vstupního souboru na minimum)

# Verifikace

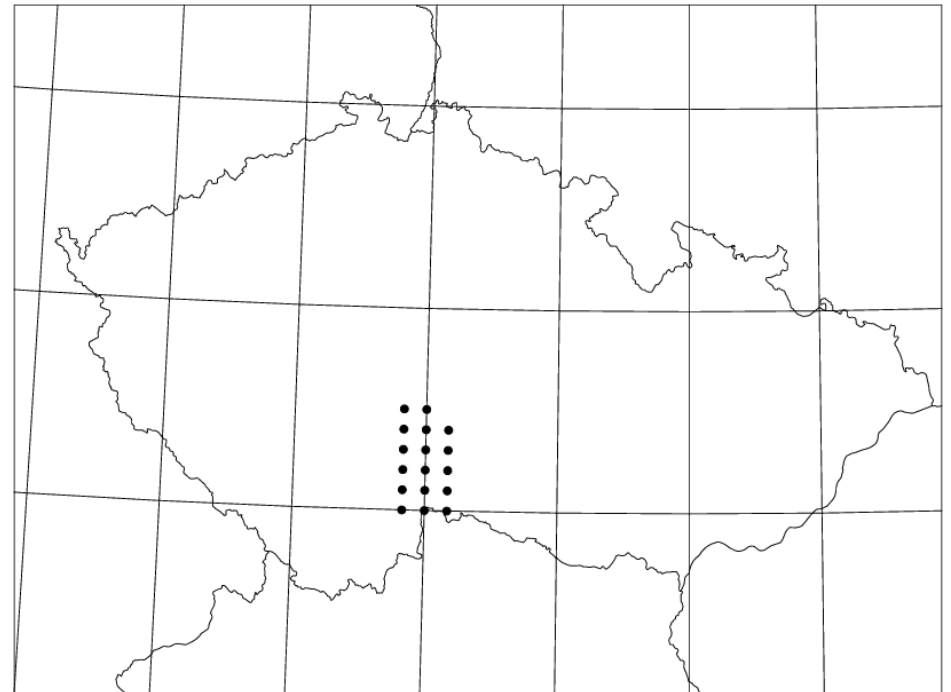
(ukázka pro hydrologické povodí LASEŇ, pravděpodobnostní předpověď a podmíněná kvantitativní předpověď srážkových úhrnů větších než zvolený práh = 0.01mm/6hodin, délka učebního souboru od 02/2003-06/2005, délka testovací vzorky 07/2005-09/2006 )

- **subjektivní hodnocení uživatele**

- **objektivní metody:**

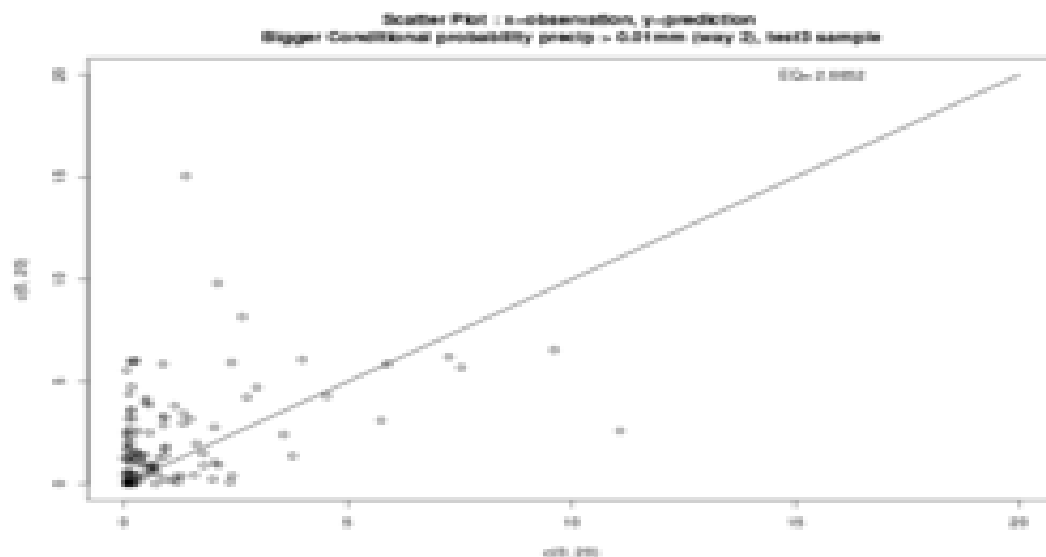
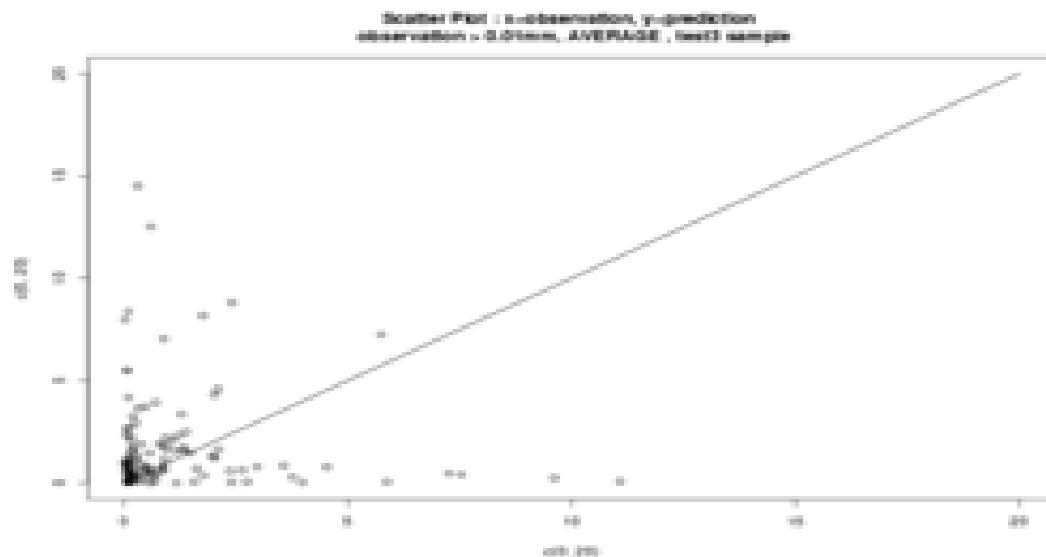
- „scatter plots“
- kvadratická chyba
- „Brier score“ a jejich dekompozice
- „ROC diagrams“
- „Reliability diagrams“

Mask =xxxx\_k



# „scatter plots“

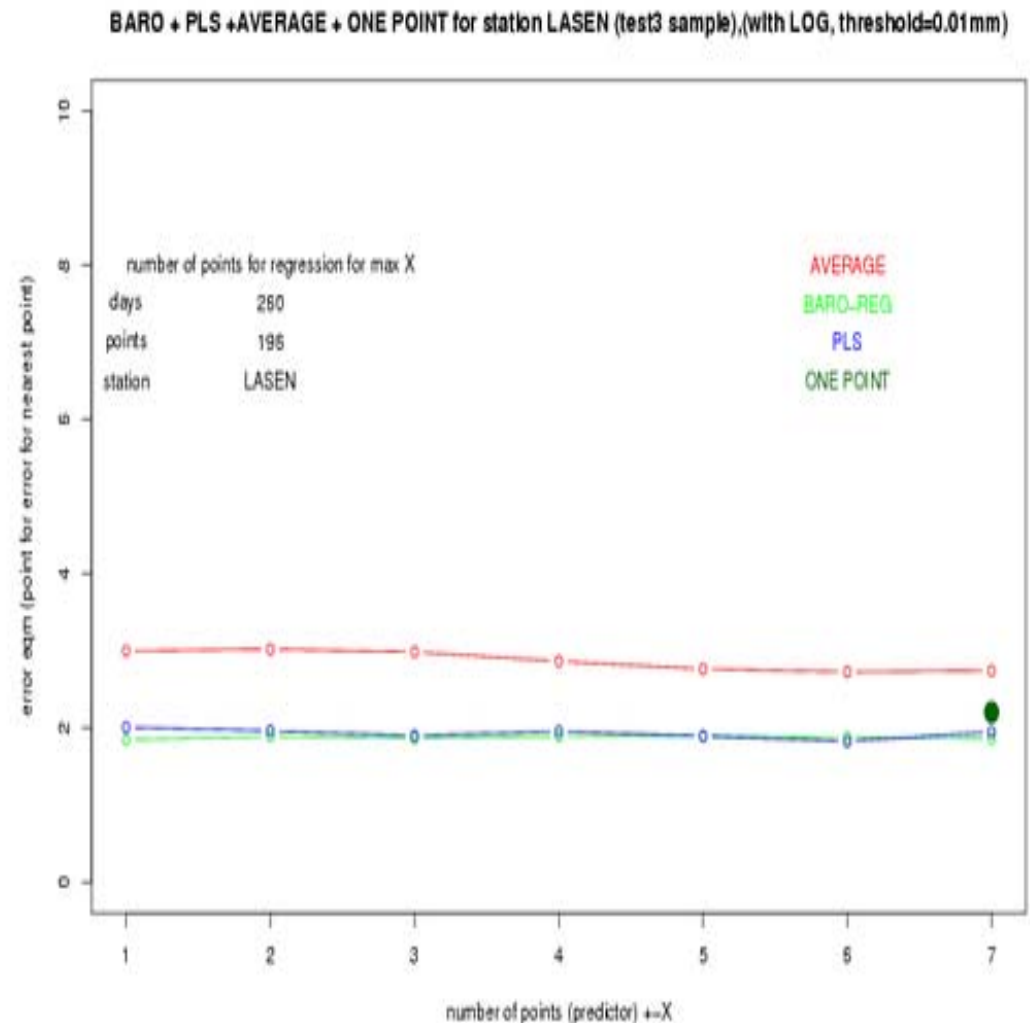
- osa  $x$  – měření, osa  $y$  – předpověď
- horní obrázek: průměr 6x6 bodů, kvadratická chyba 3.08, spodní obrázek: BMA ze 6x6 bodů, kvadratická chyba 2.66
- čím víc bodů leží u diagonály, tím je předpověď lepší





# kvadratická chyba

- kombinace Logistických regresí s Barisotropickými regresemi a PLSR, podmíněná kvantitativní předpověď
- na ose  $x$  je počet bodů od středu oblasti, které jsme brali v úvahu (1 = pole 2x2 = 4 body modelu, 7 = pole 14x14 = 196 bodů modelu)
- na ose  $y$  je kvadratická chyba pro průměr z bodů, Barisotropické regrese, PLSR, jeden bod uprostřed oblasti



# „Brier score“ a jejich dekompozice

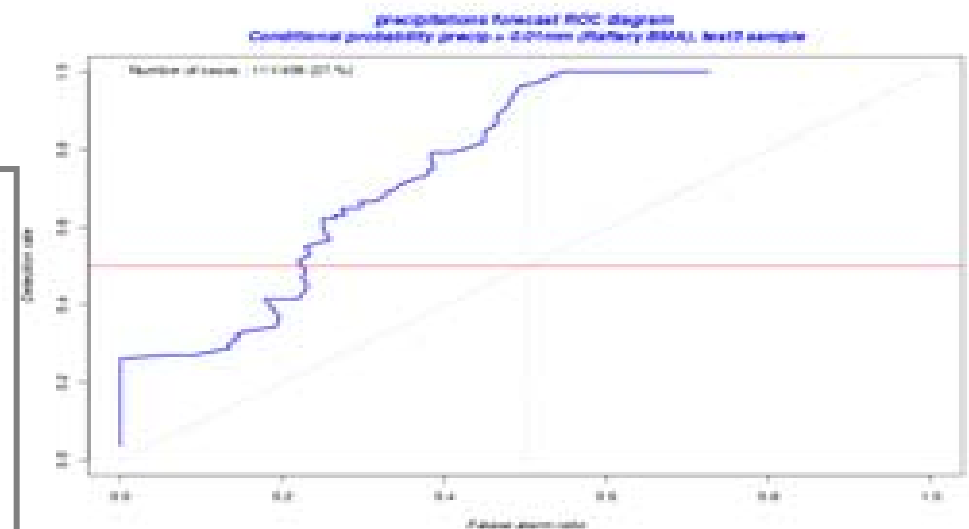
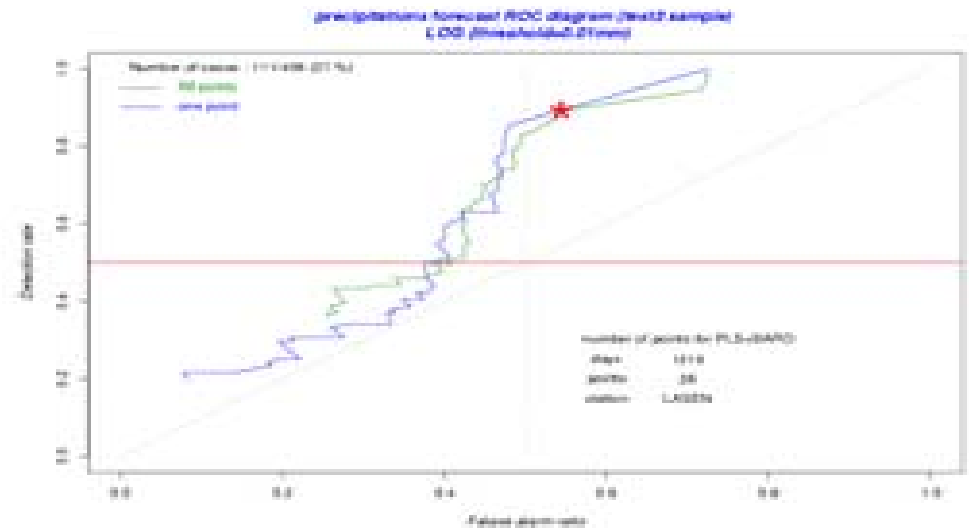
- **BRIER = REL – RES + UNC**
- **„Reliability“** - spolehlivost, zachycuje rozdíl mezi předpovědí a průměrem měření přes celý interval pro vybranou pravděpodobnostní třídu předpovídaných hodnot
- **„Resolution“** - rozlišení, zachycuje rozdíl mezi průměrem měření a průměrem měření pro vybranou pravděpodobnostní třídu předpovídaných hodnot
- **„Uncertainty“** - náhodnost, odpovídá rozptylu měření, zachycuje, jak obtížné je předpovídat

- analog ke kvadratické chybě pro pravděpodobnostní předpověď, zachycují těsnost předpovědi, čím nižší, tím lépe
- čím je RES vyšší než REL, tím líp

	LOG	BMA
BRIER	0.1641	0.1272
REL	0.0227	0.0061
RES	0.0572	0.0776
UNC	0.1987	0.1987

# „pseudo-ROC diagrams“

- transformace předpovědi s různou pravděpodobností do indexů
- osa x: FAR =  $b / (b+d)$
- osa y: DR =  $d / (c+d)$
- čím blíže k levému hornímu rohu, tím lépe
- horní obrázek: Logistické regrese, spodní obrázek „BMA“



	prediction	
observation	no	yes
no	a	b
yes	c	d

# „Reliability diagrams“

- osa x – pravděpodobnost předpovědi v procentech
- osa y – pozorované četnosti
- čím více bodů leží u diagonály, tím lépe
- horní obrázek: Logistické regrese, dolní obrázek: BMA

