

# Simulace

## Simulace dat

Menu:	QCExpert	Simulace	Simulace dat
-------	----------	----------	--------------

Tento modul je určen pro generování pseudonáhodných dat s danými statistickými vlastnostmi. Nabízí čtyři typy rozdělení: normální, logaritmickeo-normální, rovnoměrné a obecné rozdělení lambda. Pro každé rozdělení je možné zadat hodnoty parametrů jako střední hodnota, rozptyl, šikmost. Navíc je možné simulovat náhodný výskyt vybočujících hodnot s pravděpodobností 0 až 1 a autokorelaci dat prvního řádu s autokorelačním koeficientem -1 až 1. Vygenerovaná data lze použít pro simulaci reálných procesů a ověření postupů analýzy předem, bez nutnosti disponovat dostatkem dat. Dat s rovnoměrným rozdělením lze použít jako tabulku náhodných čísel pro výběr vzorků při kontrole nebo statistické přejímce s vracením.

## Parametry

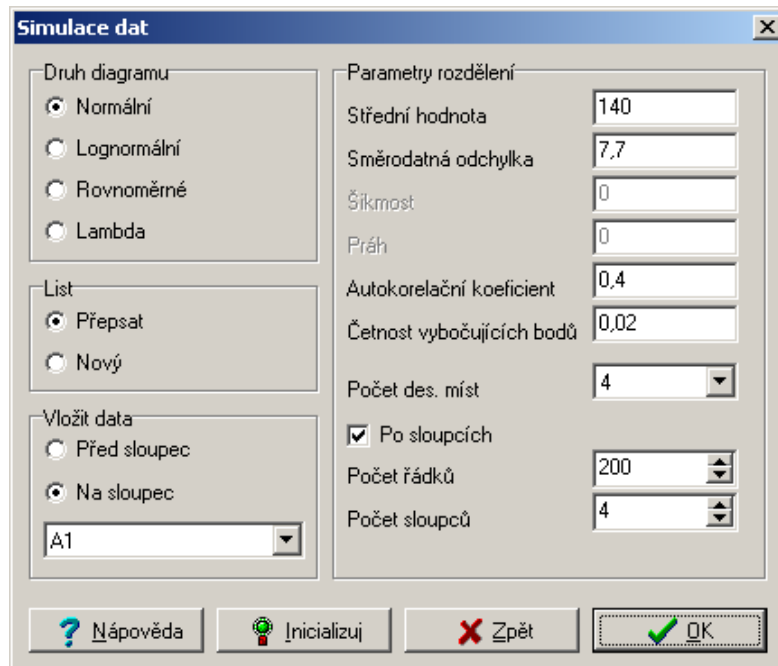
Tato metoda nemá vstupní data. Parametry pro jednotlivá rozdělení se zadávají v panelu *Simulace dat*, Obrázek 1. Pro některá rozdělení lze zadat jen určité parametry, jak uvádí následující tabulka:

Tabulka 1 Použitelnost parametrů pro jednotlivá rozdělení

	Normální	Lognormální	Rovnoměrné	Lambda
Střední hodnota	✓	✓	✓	✓
Směrodatná odchylka	✓ kladná	✓ kladná	✓ kladná	✓ kladná
Šikmost	✗	✗	✓	✓
Práh	✗	✓	✗	✗
Autokorelační koeficient	✓ -1 až 1	✓ -1 až 1	✓ -1 až 1	✓ -1 až 1
Četnost vybočujících bodů	✓ 0 až 1	✓ 0 až 1	✓ 0 až 1	✓ 0 až 1

✗ - parametr nelze použít      ✓ - parametr lze použít s případným omezením

Vygenerovaná data jsou náhodná a mohou splňovat nastavené parametry pouze přibližně. Nenulová šikmost u rovnoměrného rozdělení způsobí jeho změnu na rozdělení lichoběžníkové. Směrodatná odchylka u rovnoměrného rozdělení má význam rozpětí. Zadání nenulového autokorelačního koeficientu ovlivní výsledné parametry rozdělení (například zvýší směrodatnou odchylku).



Obrázek 1 Panel pro zadání parametrů simulace dat

## Šíření chyb

Menu:	QCExpert	Simulace	Šíření chyb
-------	----------	----------	-------------

Tento modul je určen pro případy, kdy je třeba znát statistické chování veličiny, kterou neměříme přímo, ale počítáme ji z jiných naměřených hodnot, nebo z hodnot, jejichž statistické chování známe. Pomocí tohoto modulu lze určit rozsah výsledných hodnot, střední hodnotu, interval spolehlivosti a také absolutní a relativní příspěvky jednotlivých proměnných k celkové variabilitě. Použitá metoda se někdy nazývá také výpočet nejistot, nebo propagace chyb. K výpočtu vlastností dané funkce náhodných proměnných je použito jednak metody Monte Carlo, jednak Taylorova rozvoje do druhého stupně.

## Data a parametry

Základním vstupem této metody jsou náhodné proměnné a jejich funkce. Náhodné proměnné mohou být zadány:

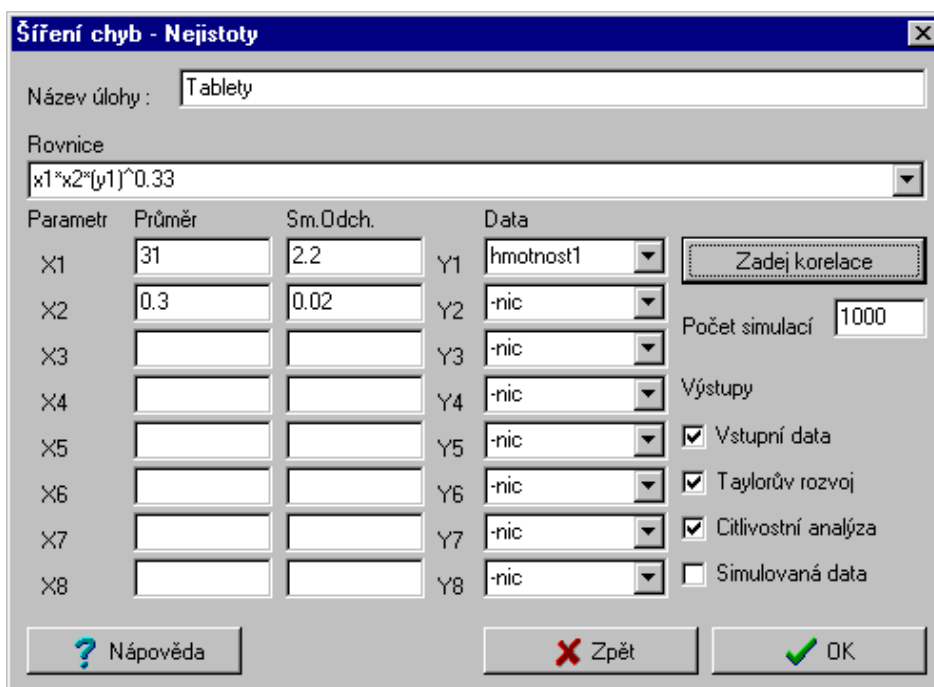
a) Pomocí střední hodnoty a směrodatné odchylky. Taková proměnná je považována za proměnnou s normálním rozdělením a při simulaci se používá generátoru náhodných čísel s normálním rozdělením. Tyto proměnné jsou v programu označovány symbolem  $X$  nebo  $x$  ( $x_1, x_2, \text{atd.}$ ).

b) Pomocí vzorku skutečných naměřených dat. U takové proměnné je při simulaci respektováno její skutečné rozdělení, které nemusí být normální. Je použito techniky vzorkování (resamplingu) podle empirické (výběrové) distribuční funkce. Tento způsob zadání proměnné se doporučuje tehdy, máme-li k dispozici dostatečný počet (alespoň 20) reálných dat, zvláště nejsme-li si jisti jejich rozdělením. Tyto proměnné jsou v programu označovány symbolem  $Y$  nebo  $y$  ( $y_1, y_2, \text{atd.}$ ) a jsou zadány obvyklým způsobem ve sloupcích.

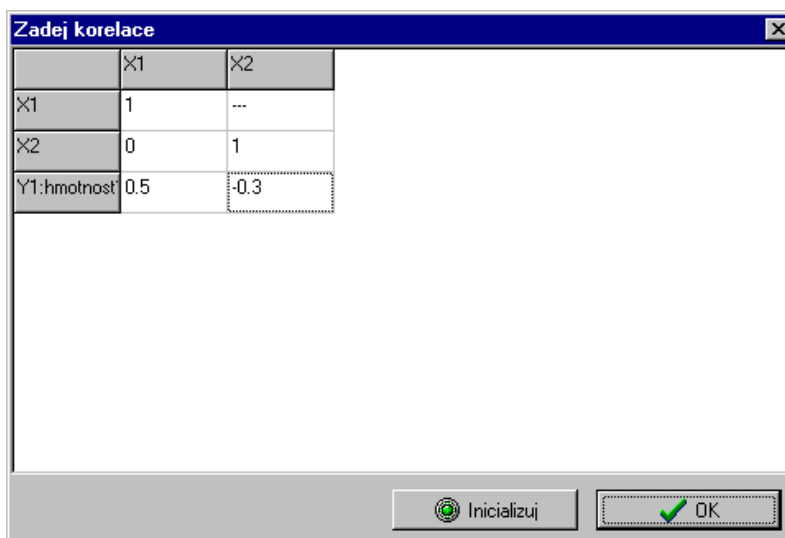
Oba typy proměnných lze libovolně kombinovat. Maximální počet proměnných  $X$  a  $Y$  je 20 (10+10). V dialogovém panelu se proměnné (parametry)  $x_i$  definují pomocí průměru a směrodatné odchylky (je nutno zadat obojí), proměnné  $y_i$  se definují pomocí příslušného sloupce dat v aktuálním listu. Do okénka *Rovnice* se zapíše matematický pomocí proměnných  $x_1, x_2, \dots, y_1, y_2, \dots$  a obvyklé syntaxe. Například  $5 * x_1 * (1 - x_2) / \log(3.124 * y_1)$ . Dále je před výpočtem možno zadat párové korelační koeficienty mezi proměnnými  $x_i$  a  $x_j$ , popř.  $x_i$  a  $y_j$ , jsou-li známy. Párové korelace mezi proměnnými  $y_i$  a  $y_j$  se vypočítají automaticky z dat. Hodnota korelačních koeficientů může značně ovlivnit odhady střední hodnoty, rozptylu, intervalů spolehlivosti a kvantilů metodou Taylorova rozvoje. Tlačítko *Inicializuj* v panelu pro zadání korelací vynuluje všechny koeficienty. Pokud se

korelace nezadají, jsou považovány za nulové. Hodnoty korelačních koeficientů musí být v intervalu  $(-1, 1)$ .

Doporučený počet simulací je pro většinu úloh 100 až 1000. Zaškrtnuté položky výstupů budou zahrnuty do protokolu. Simulovaná data (jsou-li požadována) se uloží do zvláštního listu v okně *Protokol*. Těchto dat lze využít k podrobnější analýze předpokládaného rozdělení výsledné veličiny pomocí modulu Základní statistika.



Obrázek 2 Zadání podmínek výpočtu pro šíření chyb



Obrázek 3 Zadání předpokládaných korelací pro šíření chyb

## Protokol

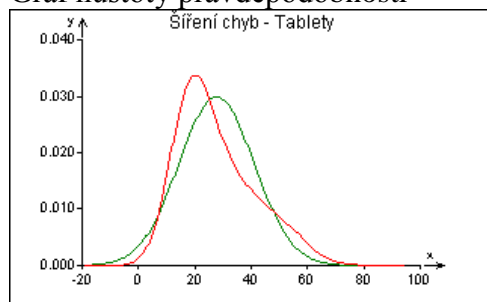
<b>Funkce</b>	Zadaný tvar vztahu pro výslednou proměnnou.
<b>Vstupní veličiny</b>	Charakteristiky proměnných X zadaných pomocí střední hodnoty a směrodatné odchylky.
Střední hodnota	Zadaná střední hodnota.
Sm. odchylka	Zadaná směrodatná odchylka.
95% interval	Vypočítaný 2,5% a 97,5% kvantil dat.

+3sigma	Interval, v němž je 99,73% dat.
<b>Vstupní data</b>	Charakteristiky proměnných Y zadaných pomocí dat.
Střední hodnota	Aritmetický průměr dat, odhad střední hodnoty.
Sm. odchylka	Vypočítaná směrodatná odchylka.
95% interval	Vypočítaný 2,5% a 97,5% kvantil dat.
+3sigma	Interval, v němž je 99,73% dat.
<b>Výsledná hodnota</b>	Charakteristiky výsledné proměnné získané simulací Monte Carlo.
Medián	Medián je spolehlivějším odhadem střední hodnoty než aritmetický průměr u asymetrických dat, nebo dat s velkou špičatostí.
Střední hodnota	Odhad střední hodnoty aritmetickým průměrem.
Sm. odchylka	Odhad směrodatné odchylky.
95% interval	2,5% a 97,5% kvantil simulovaných dat.
+3sigma	Interval, v němž je 99,73% dat.
<b>Interval výsledných hodnot</b>	Nejmenší a největší výsledná hodnota vygenerovaná metodou Monte Carlo.
<b>Citlivostní analýza</b>	
Absolutní citlivost	Absolutní vliv jednotlivých proměnných na výslednou veličinu. Počítá se jako parciální derivace dané funkce podle každé proměnné v bodě jejího průměru. Je to tedy citlivost výsledné veličiny na (malou) jednotkovou změnu dané proměnné. Z hlediska variability lze interpretovat jako očekávané zvýšení směrodatné odchylky výsledné veličiny při jednotkové změně směrodatné odchylky dané proměnné.
Relativní citlivost	Vliv jednotlivých proměnných na výslednou veličinu vzhledem k jejich skutečné variabilitě. Počítá se jako absolutní citlivost násobená směrodatnou odchylkou příslušné proměnné. Je-li relativní citlivost pro určitou proměnnou nízká proti ostatním, nemá příliš smysl se snažit o snížení její variability, resp. její statistickou regulaci. Je-li naopak vysoká, přinese regulace, nebo snížení variability příslušné proměnné výrazné snížení variability výsledné veličiny
<b>Aproximace metodou Taylorova rozvoje</b>	Alternativní metoda výpočtu pomocí analytického Taylorova rozvoje do druhého řádu. Tato metoda bere v úvahu i kovariance (resp. korelace) mezi proměnnými.
Prostý průměr	<i>Nesprávná (!)</i> střední hodnota získaná pouhým dosazením středních hodnot proměnných do funkce.
Opravený průměr	Odhad střední hodnoty metodou Taylorova rozvoje bez uvažování kovariančního členu.
Průměr s kovariancí	Odhad střední hodnoty metodou Taylorova rozvoje s uvažováním kovariančního členu.
Opravená sm. odchylka	Odhad směrodatné odchylky metodou Taylorova rozvoje bez uvažování kovariančního členu.
Sm. Odch. s kovariancí	Odhad směrodatné odchylky metodou Taylorova rozvoje s uvažováním kovariančního členu.
95% interval	95% interval dat, tedy 2.5% a 97.5% kvantil. Vypočítaný za předpokladu normality rozdělení výsledné veličiny!
Interval +3sigma	ca. 99.73% interval dat, tedy asi 0.135% a 99.865% kvantil. Vypočítaný za předpokladu normality rozdělení výsledné veličiny!
Simulovaná data	Ukládají se do zvláštního listu. Data představují simulaci očekávaných hodnot výsledné veličiny.

## Grafy

Grafy jsou generovány pouze v případě použití kvadratické metody.

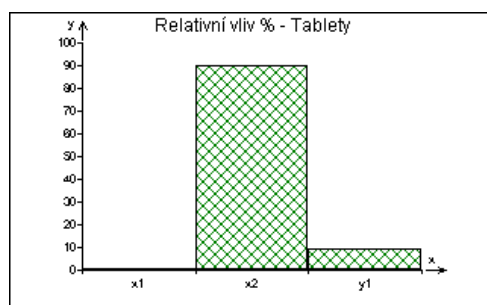
### Graf hustoty pravděpodobnosti



Porovnání předpokládaného skutečného rozdělení dat (červeně) s rozdělením normálním (zeleně). Zřetelný rozdíl upozorňuje na porušení normality.

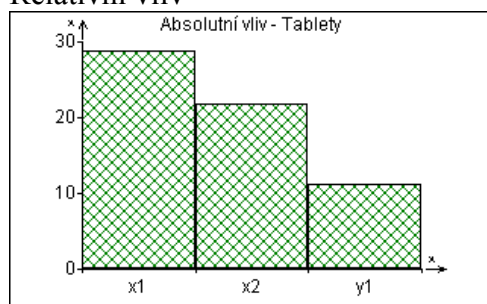
### Absolutní vliv %

Procentové vyjádření absolutní citlivosti, viz *Protokol*.



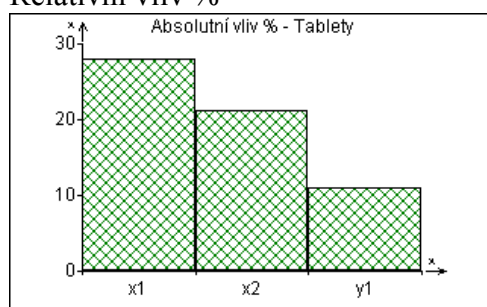
### Relativní vliv

Relativní citlivost, viz *Protokol*.



### Relativní vliv %

Relativní citlivost vyjádřená v procentech viz *Protokol*.



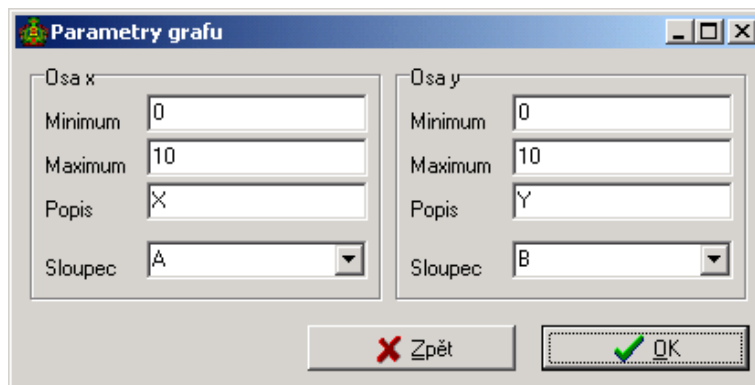
## Grafická simulace dat

Menu:

Tento modul umožňuje interaktivní simulaci dat pomocí myši přímo do prázdného grafu se souřadnicemi  $x$ ,  $y$ . Data umístěná do grafu se pak přenesou do určených sloupců v datové tabulce. Grafická simulace je účinný způsob rychlého získání dat předem známého charakteru, která pak mohou být například použita k ověření statistických postupů „na nečisto“ apod.

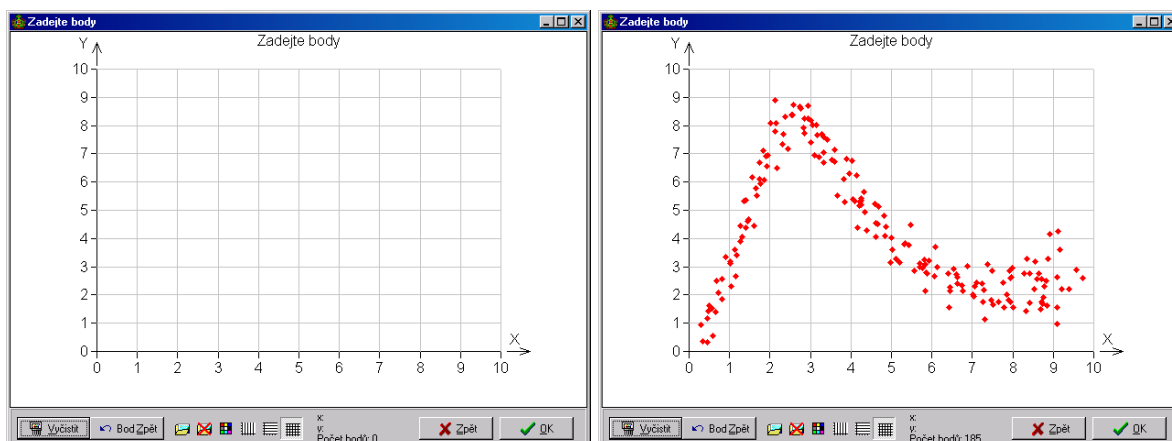
## Parametry

V dialogovém okně (Obrázek 4) se v polích *Minimum* a *Maximum* nastaví rozsah osy *X* a osy *Y* v grafu, může se zvolit název záhlaví sloupce v datové tabulce v polích *Popis*. V položce *Sloupec* se vybere sloupec v datové tabulce, kam se zadaná data vloží. Data se vkládají do tabulky vždy od prvního sloupce, eventuální obsah sloupců se přepíše. Stiskem tlačítka *OK* se otevře okno s prázdným grafem s vyznačenými osami *X* a *Y*.

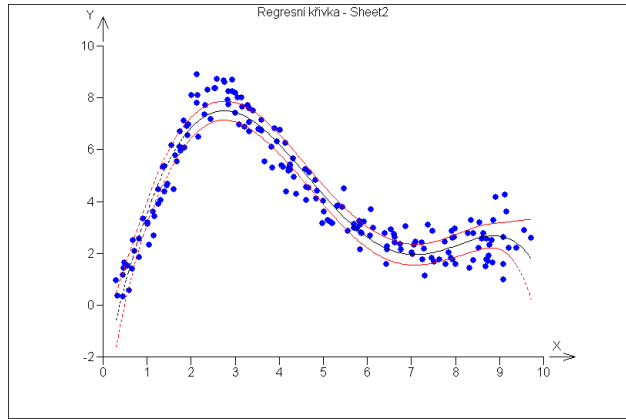


Obrázek 4 Dialogové okno pro grafickou simulaci

V prázdném grafu (Obrázek 5 A) lze myší zadávat body, například předpokládané naměřené hodnoty veličin, časové průběhy, závislosti, a podobně (Obrázek 5 B). Pro lepší orientaci lze v grafu zobrazit síť. Stiskem tlačítka *Vyčistit* lze data smazat a začít znova. Po stisku *OK* se zadaná data přenesou do zvolených sloupců datové tabulky. V pravé dolní části grafického okna se zobrazují aktuální souřadnice polohy ukazatele myši v grafu a počet dosud zadaných bodů.

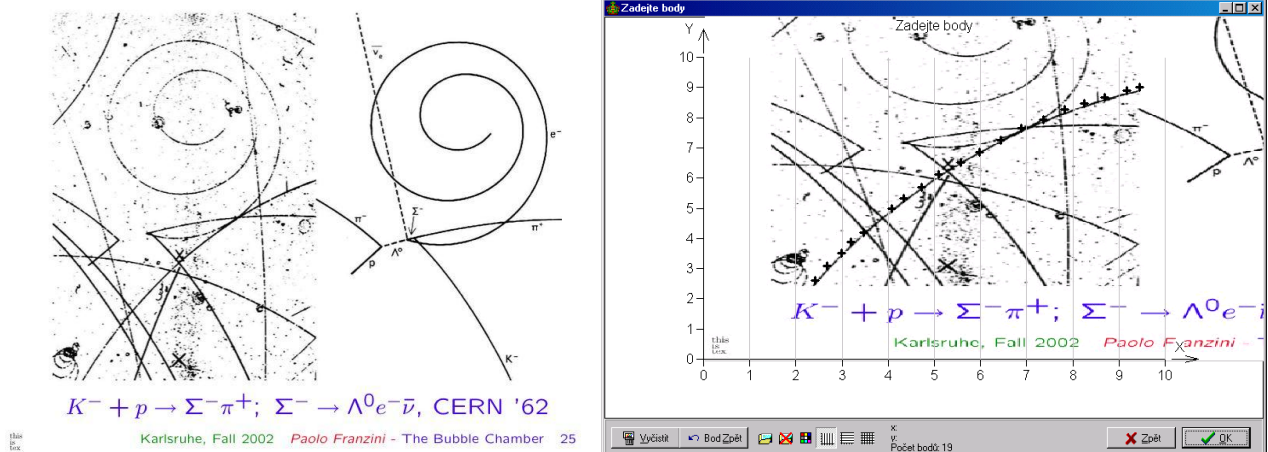


Obrázek 5 Okno pro zadání dat: prázdné (A) a se zadanými body (B)



**Obrázek 6 Příklad zpracování zadaných dat lineární regresí**

**Digitalizace obrazu:** Tlačítkem *Vložit obrázek* lze na plochu se souřadnicemi umístit bitovou mapu (v grafickém formátu GIF, JPG, BMP, ICO, EMF, nebo WMF) a tvary (např. trajektorie, profily, okraje, průhyby, konstrukční prvky, fotokopie záznamů ze zapisovačů, oskenované grafy, atd.) převést pomocí myši na body v souřadnicové síti. Po vložení lze obrázek posouvat a měnit jeho velikost kliknutím pravým tlačítkem myši. Po nastavení pozice a velikosti obrázku přejdeme k zadávání bodů kliknutím levým tlačítkem myši v okně grafu mimo obrázek.



**Obrázek 7 Digitalizace bitového obrazu**