VENTANA systems, inc.

Vensim® Software

Linking systems thinking to powerful dynamic models

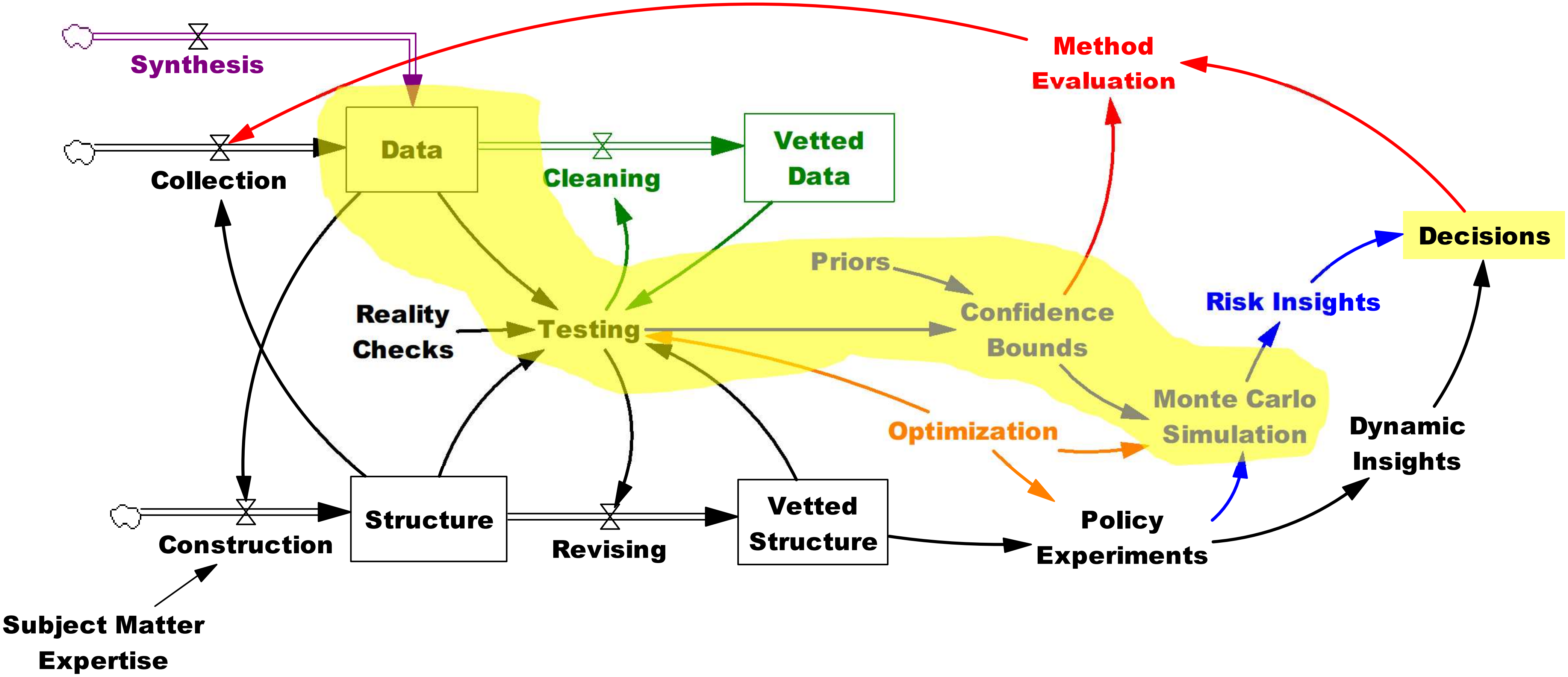
Calibration with Vensim

Tom Fiddaman
2022

Agenda

- Introduction
- Synthetic data
- Naïve calibration
- Maximum likelihood
- Markov Chain Monte Carlo (MCMC)
- Kalman filtering (briefly)

The Big Picture



Overview

- **Lotka-Volterra Predator-Prey model**

- **Generate synthetic data, by adding:**
  - Measurement error to the stocks of elk and wolves
  - Driving noise to the flows of births and mortality

- **Estimate parameters of the model from the data, by various methods**
  - Optionally, use mismatched structure (2nd order data-generating model, first order estimated model)

Caveats

In order to get done, we're approaching this problem a bit fast and loose. Be aware:

- **There is structural uncertainty as well as parameter uncertainty**
- **Statistics deserve deeper thought**
  - Weights
  - Covariance
  - Autocorrelation
  - Distributional assumptions
  - Measurement error & driving noise (Kalman filter)
- **We should be testing for multiple optima with multistart calibration runs**
- **Sample sizes for sensitivity and MCMC may be too small**

Other Pitfalls

- State dependent noise
- Sample size
- Data quality
- Autocorrelated errors
- Error covariance
- Measurement error
- State estimation
- Endogeneity

Calibration

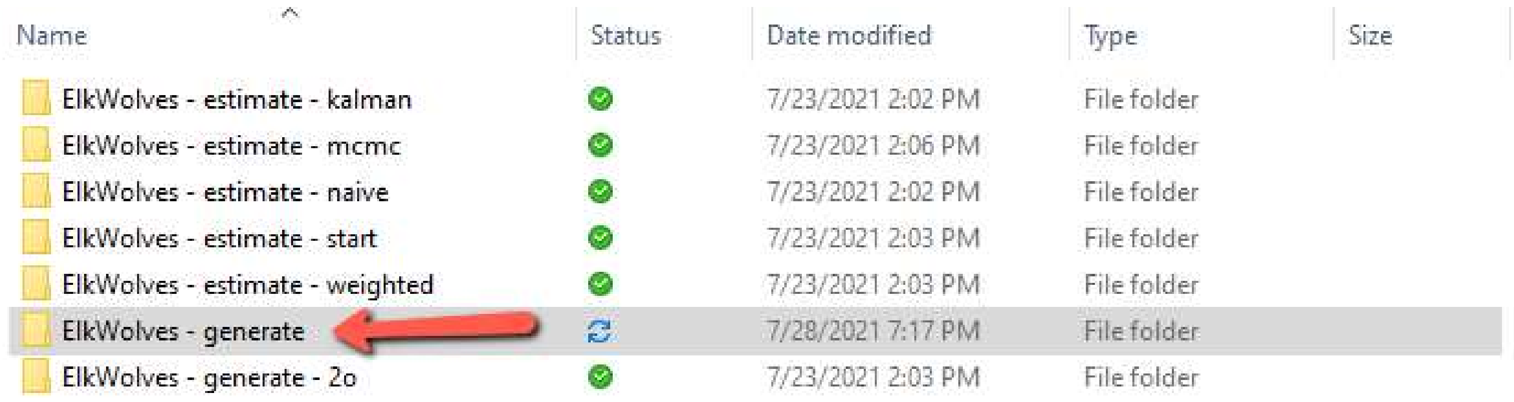
- **Purposes**
  - Make better predictions or measurements
  - Reject models that can't replicate data (potentially a weak test of quality)
  - Learn about the model
  - Learn about the data
  - Provide face validity for reviewers
- **Closeness Measures**
  - Sum of squared errors & R^2
  - Mean Absolute Deviation
  - Mean Absolute Percent Error
  - Log Likelihood

- **Process**
  - Assume the model structure is right
    - If possible, test alternatives!
  - Simulate the model
    - Measure the closeness
    - Adjust the constants in the model
    - Iterate to improve
  - After convergence, evaluate the fit
    - Decide if the model needs revision
    - Investigate puzzles in the data

Model Tour

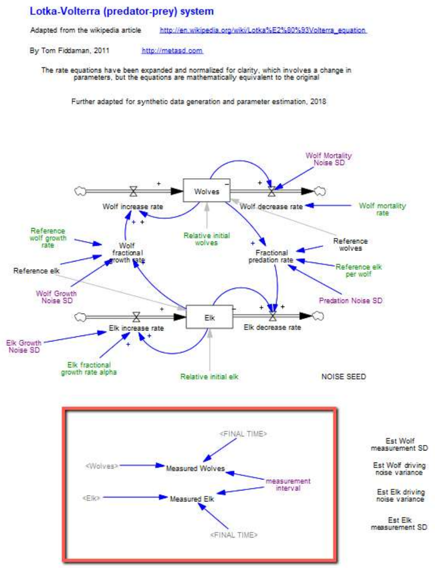
Synthetic Data Generation

- **Open the data-generating model**



Name	Status	Date modified	Type	Size
ElkWolves - estimate - kalman	✓	7/23/2021 2:02 PM	File folder	
ElkWolves - estimate - mcmc	✓	7/23/2021 2:06 PM	File folder	
ElkWolves - estimate - naive	✓	7/23/2021 2:02 PM	File folder	
ElkWolves - estimate - start	✓	7/23/2021 2:03 PM	File folder	
ElkWolves - estimate - weighted	✓	7/23/2021 2:03 PM	File folder	
ElkWolves - generate	⟳	7/28/2021 7:17 PM	File folder	
ElkWolves - generate - 2o	✓	7/23/2021 2:03 PM	File folder	

- **Run, and take a look at the "measured" variables (red box)**



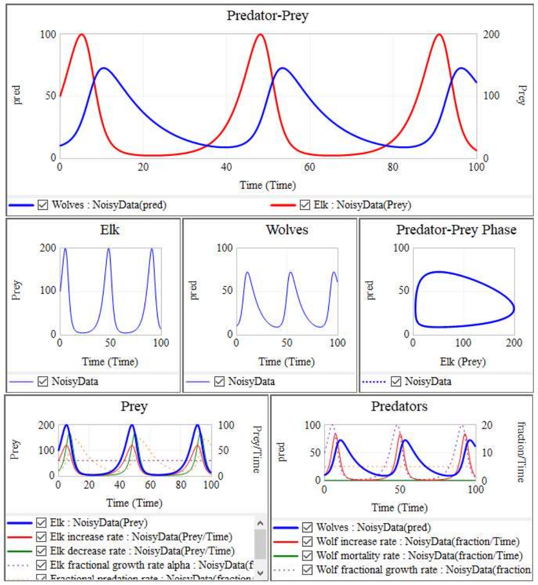
Lotka-Volterra (predator-prey) system

Adapted from the wikipedia article http://en.wikipedia.org/wiki/Lotka%E2%80%93Volterra_equation

By Tom Fiddaman, 2011 http://metasd.com

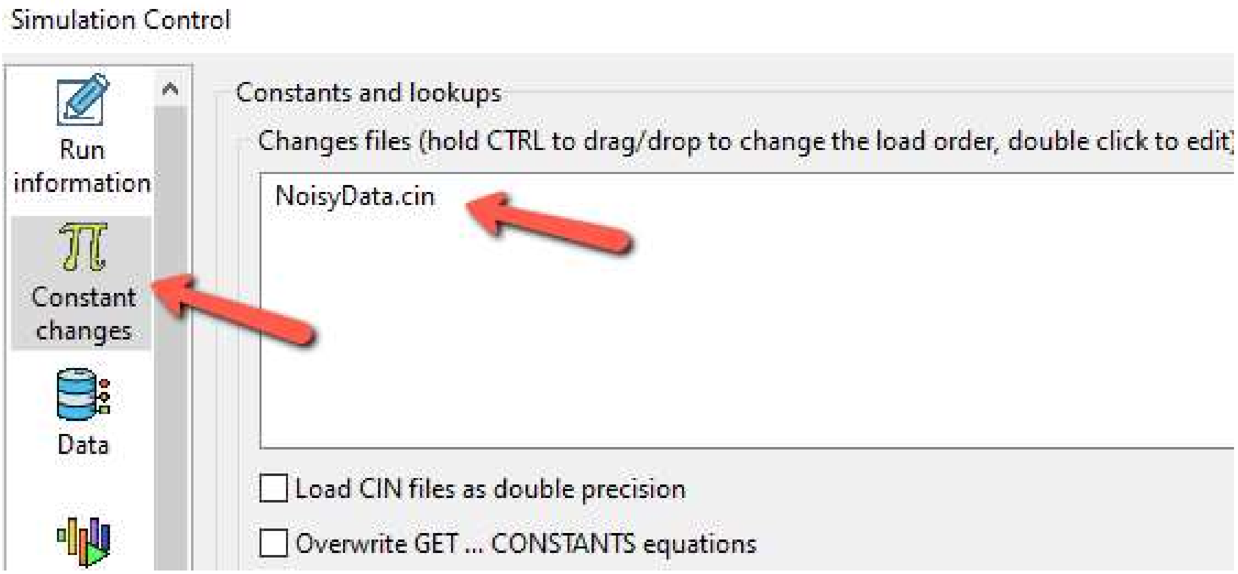
The rate equations have been expanded and normalized for clarity, which involves a change in parameters, but the equations are mathematically equivalent to the original

Further adapted for synthetic data generation and parameter estimation, 2018



Synthetic Data Generation

- **Load "NoisyData.cin" and run the model**



Constants and lookups

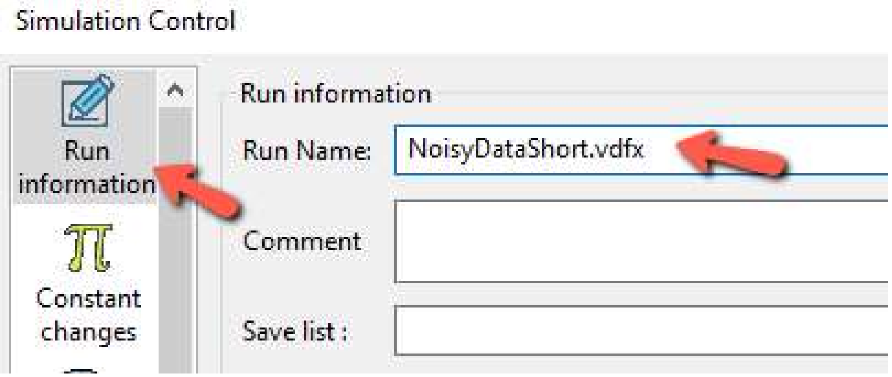
Changes files (hold CTRL to drag/drop to change the load order, double click to edit)

NoisyData.cin

□ Load CIN files as double precision

□ Overwrite GET ... CONSTANTS equations

- **Run again, but add "short" to the run name, and load "Short.cin" to set FINAL TIME=40**

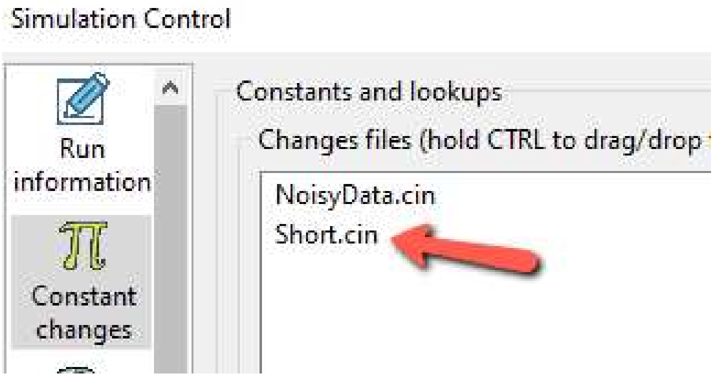


Run information

Run Name: NoisyDataShort.vdfx

Comment

Save list :



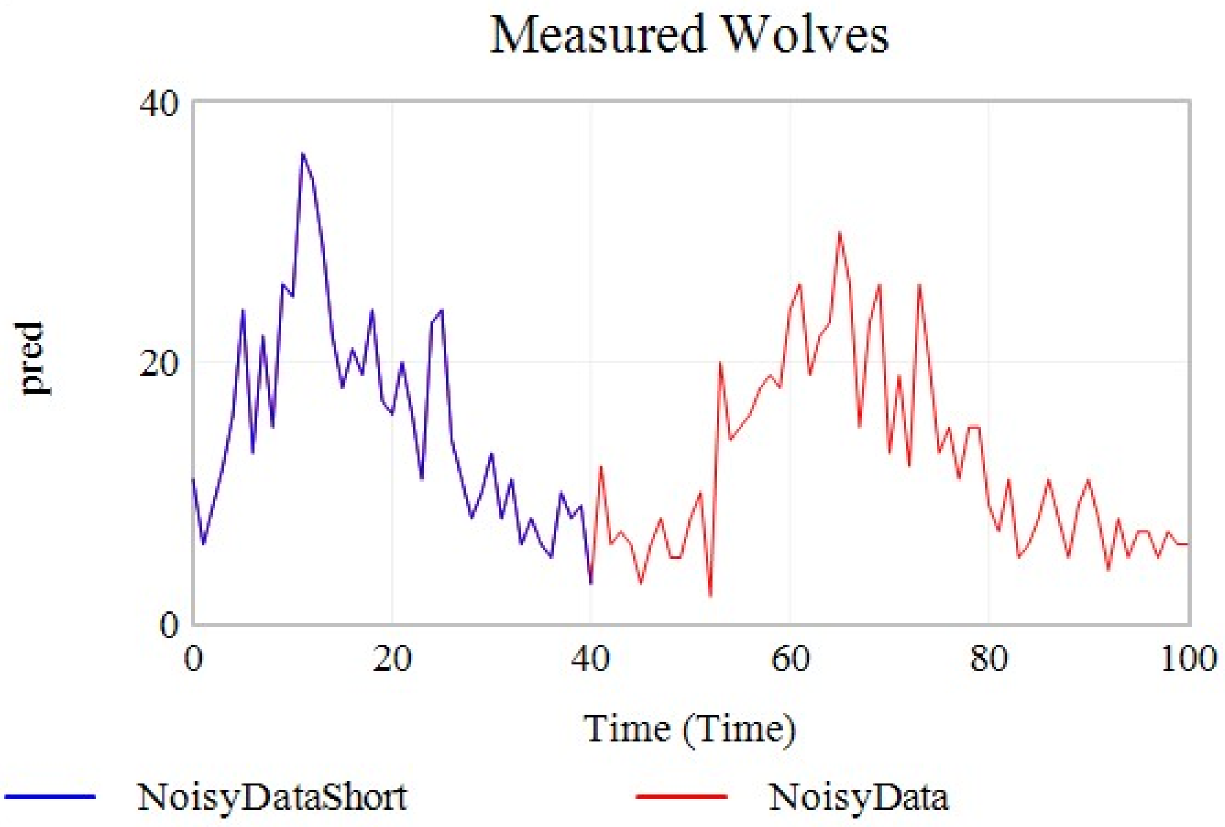
Constants and lookups

Changes files (hold CTRL to drag/drop

NoisyData.cin
Short.cin

Synthetic Data Generation

- **This gives you 2 datasets**
  - Short time series, used to estimate parameters; export and re-import using a savelist to restrict the information to the measured state only
  - Long time series, for comparison of later estimates with "truth"



Naïve Calibration

- Create a simple metric describing the distance of the model from the data

- Minimize the distance

Mechanics – What We Need

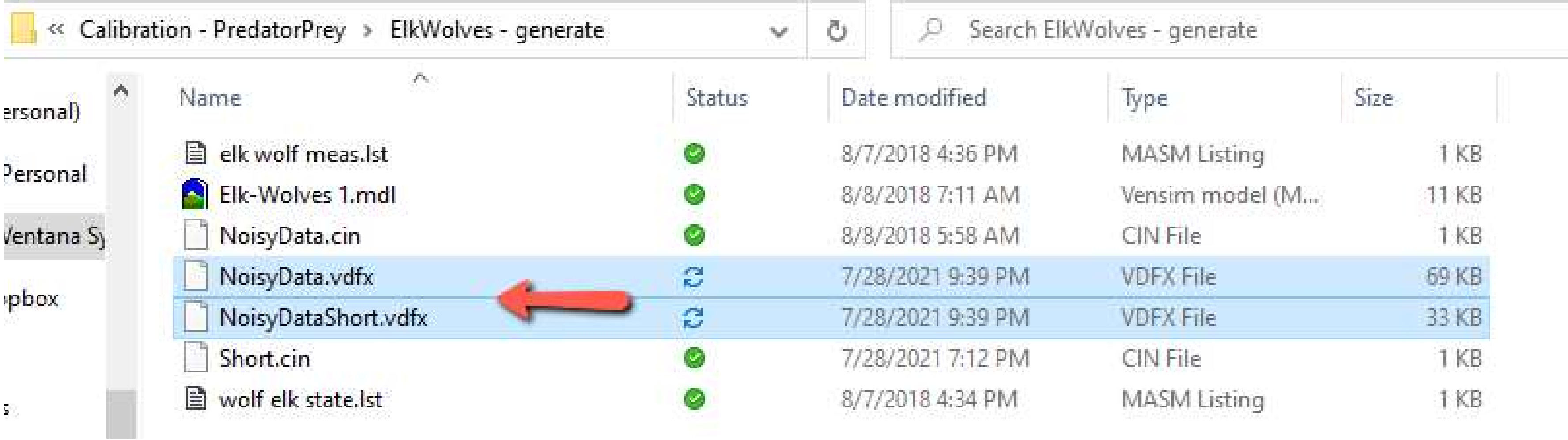
- **Data**
  - A .vdf file, or
  - Equations with GET DIRECT DATA, GET XLS DATA
  - ODBC
- **A Payoff specifying what data series to match, and how to weight each one**
- **An Optimization Control file specifying**
  - which parameters to vary, and
  - what methods to use
- **The optimizer then hill climbs to find the parameters that minimize the error between model and data**

Naïve Calibration
What do we get?

- **A run (.vdfx) with the best parameters**
- **An output file (.out) summarizing the parameters found**
  - Parameters can then be reused by loading the .out as a Changes file (like .cin files)
- **A Payoff Report (.rep) with diagnostics (optionally)**

Naïve Calibration Setup

- **Copy the data files you created from the data generator model to the “start” model folder**

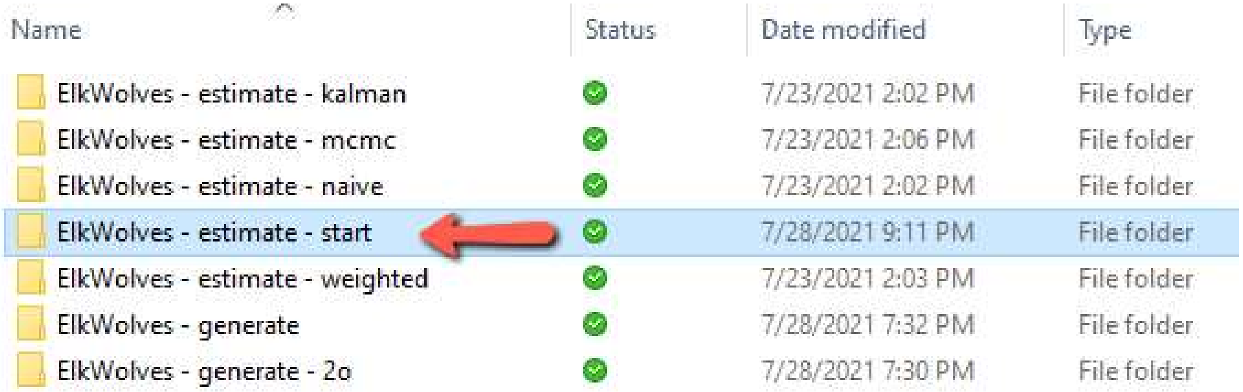


« Calibration - PredatorPrey › ElkWolves - generate

Search ElkWolves - generate

Name	Status	Date modified	Type	Size
elk wolf meas.lst	✓	8/7/2018 4:36 PM	MASM Listing	1 KB
Elk-Wolves 1.mdl	✓	8/8/2018 7:11 AM	Vensim model (M...	11 KB
NoisyData.cin	✓	8/8/2018 5:58 AM	CIN File	1 KB
NoisyData.vdfx	⟳	7/28/2021 9:39 PM	VDFX File	69 KB
NoisyDataShort.vdfx	⟳	7/28/2021 9:39 PM	VDFX File	33 KB
Short.cin	✓	7/28/2021 7:12 PM	CIN File	1 KB
wolf elk state.lst	✓	8/7/2018 4:34 PM	MASM Listing	1 KB

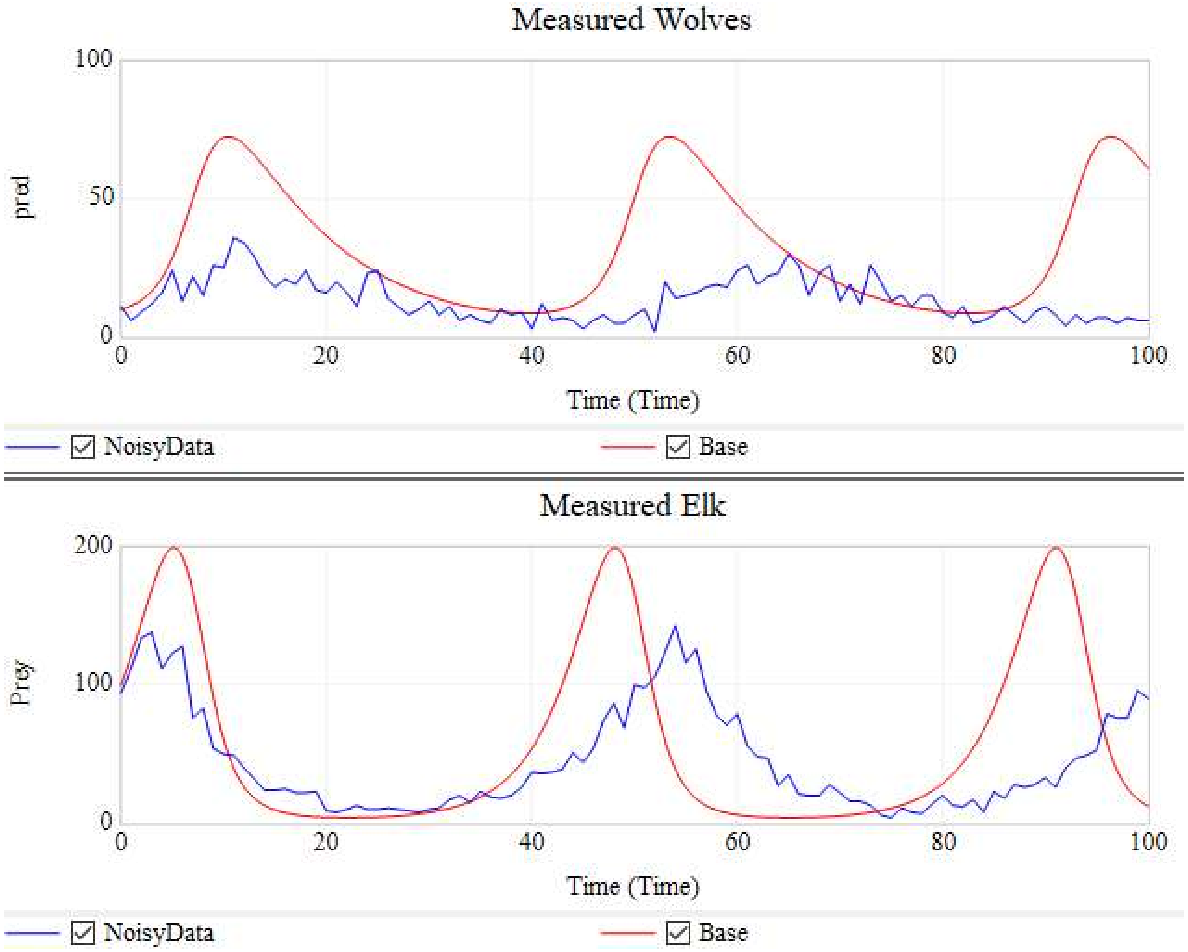
- **Open the starting point model**



Name	Status	Date modified	Type
ElkWolves - estimate - kalman	✓	7/23/2021 2:02 PM	File folder
ElkWolves - estimate - mcmc	✓	7/23/2021 2:06 PM	File folder
ElkWolves - estimate - naive	✓	7/23/2021 2:02 PM	File folder
ElkWolves - estimate - start	✓	7/28/2021 9:11 PM	File folder
ElkWolves - estimate - weighted	✓	7/23/2021 2:03 PM	File folder
ElkWolves - generate	✓	7/28/2021 7:32 PM	File folder
ElkWolves - generate - 2o	✓	7/28/2021 7:30 PM	File folder

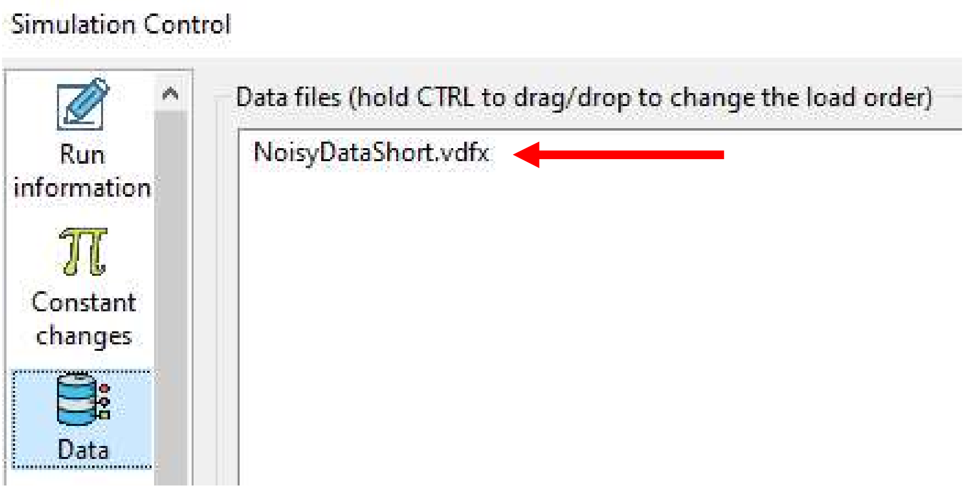
Naïve Calibration Setup

- Do a “Base” run (uncalibrated)
- Load the data from NoisyData or NoisyDataShort
- Notice how the model doesn’t fit the data well



Naïve Calibration Setup

- **Change the runname ("NaiveCal.vdfx" or similar)**
- **Go to the Data tab**
  - Load comparison data (recommend the short version)



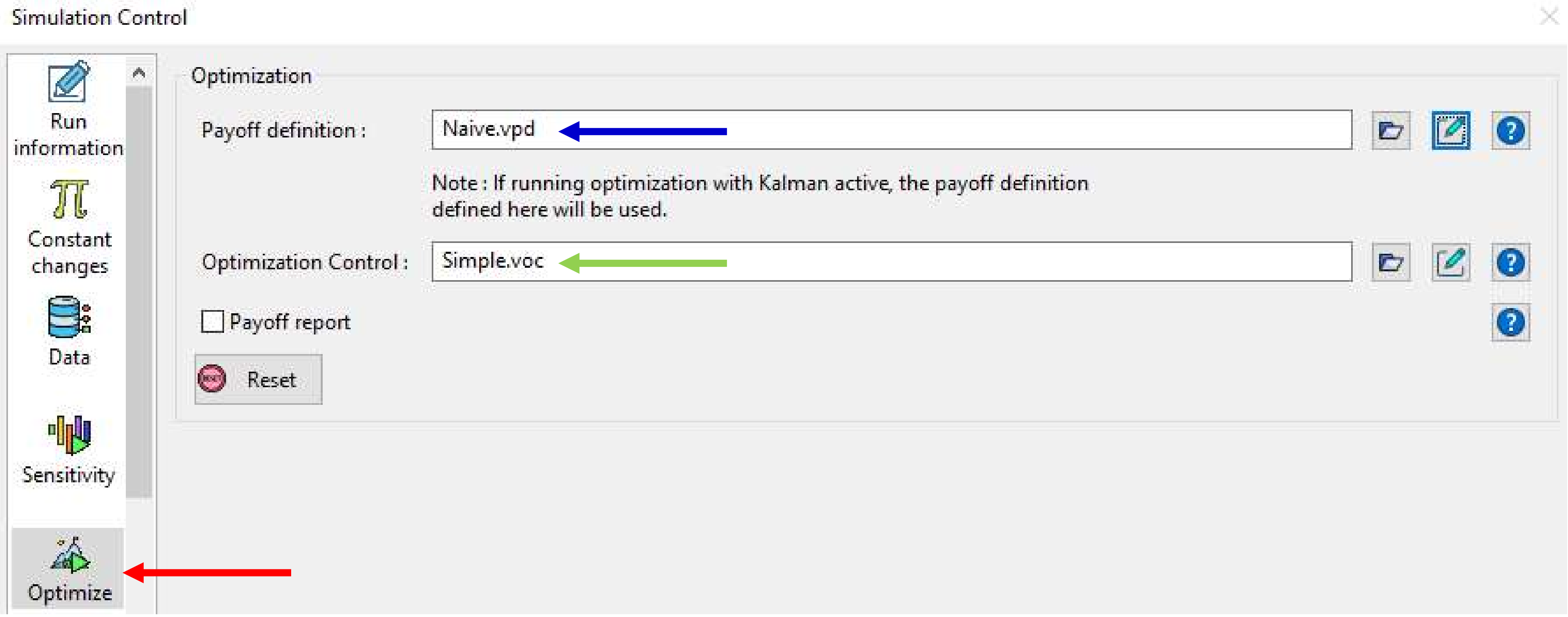
Simulation Control

Data files (hold CTRL to drag/drop to change the load order)

NoisyDataShort.vdfx

Naïve Calibration Setup (Continued)

- **Change the runname (“NaiveCal.vdfx” or similar)**
- **Go to the Optimize pane**
  - Create a Payoff (.vpd)
  - Create a Control file (.voc)



Simulation Control

Optimization

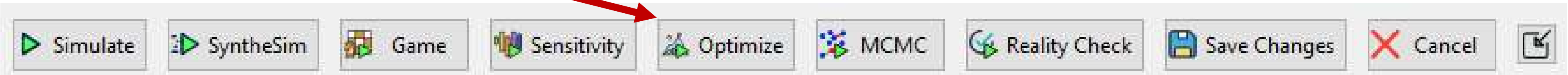
Payoff definition : Naive.vpd

Note : If running optimization with Kalman active, the payoff definition defined here will be used.

Optimization Control : Simple.voc

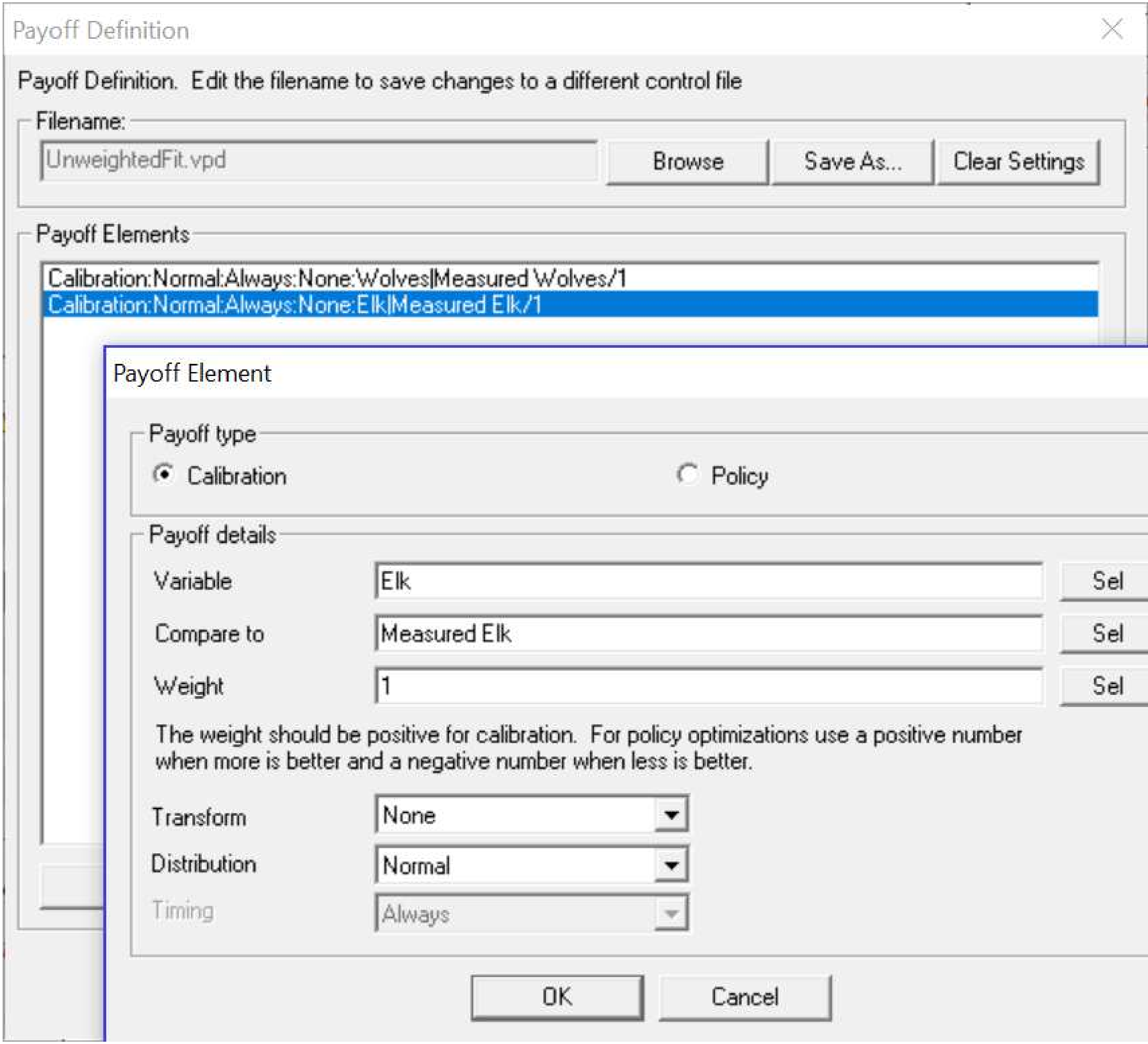
□ Payoff report

- **Hit the Optimize button**



VENTANA **Vensim®**

Payoff (.vpd)



Payoff Definition

Payoff Definition. Edit the filename to save changes to a different control file

Filename: UnweightedFit.vpd

Payoff Elements

Calibration:Normal:Always:None:Wolves|Measured Wolves/1

Calibration:Normal:Always:None:Elk|Measured Elk/1

Payoff Element

Payoff type

● Calibration ○ Policy

Payoff details

Variable: Elk

Compare to: Measured Elk

Weight: 1

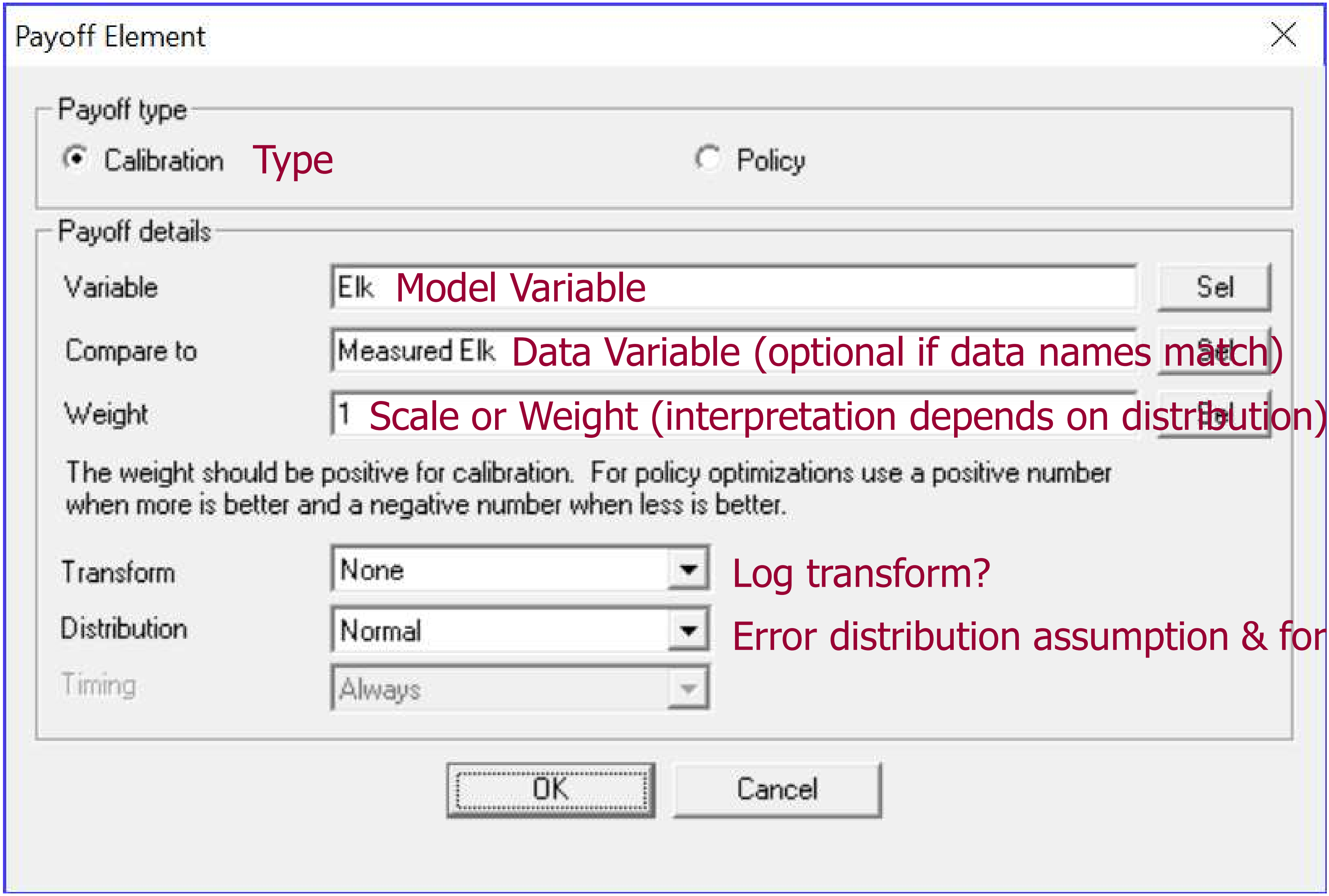
The weight should be positive for calibration. For policy optimizations use a positive number when more is better and a negative number when less is better.

Transform: None

Distribution: Normal

Timing: Always

Calibration Payoff Types



Payoff Element

Payoff type

● Calibration Type ○ Policy

Payoff details

Variable: Elk Model Variable

Compare to: Measured Elk Data Variable (optional if data names match)

Weight: 1 Scale or Weight (interpretation depends on distribution)

The weight should be positive for calibration. For policy optimizations use a positive number when more is better and a negative number when less is better.

Transform: None Log transform?

Distribution: Normal Error distribution assumption & format

Timing: Always

The Payoff File (.vpd) as text

```
*C
```

Keyword indicating type (calibration = *C, policy = *P, etc.)

```
wolves|measured wolves/1
```

Model variable | data variable / weight or scale parameter

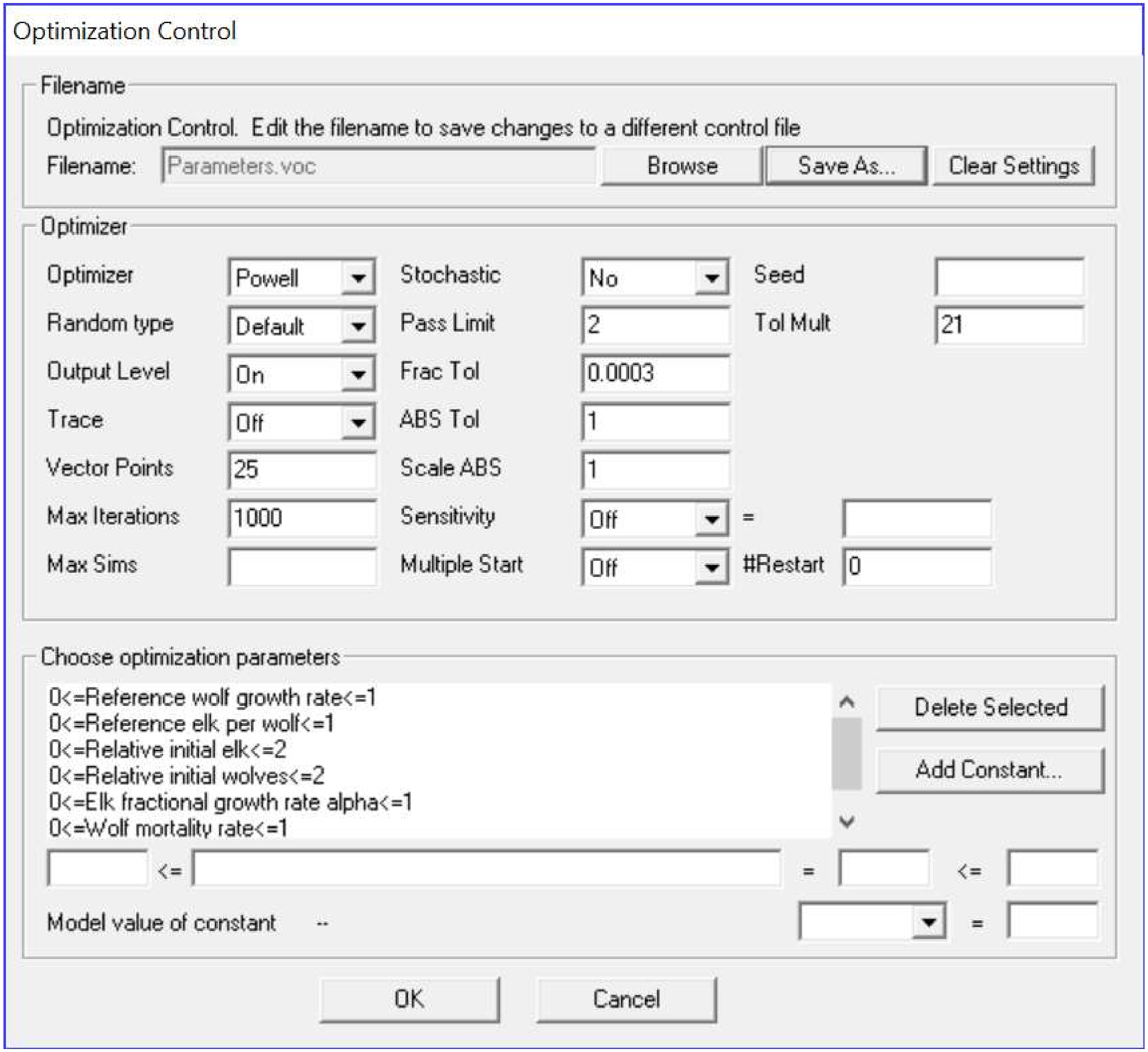
The weight can also be a variable.

Subscript ranges are OK, as long as they match.

Optimization Control File (.voc)

Method & Settings

Parameters & Bounds



Optimization Control

Filename

Optimization Control. Edit the filename to save changes to a different control file

Filename: Parameters.voc

Optimizer

Optimizer	Powell	Stochastic	No	Seed	
Random type	Default	Pass Limit	2	Tol Mult	21
Output Level	On	Frac Tol	0.0003		
Trace	Off	ABS Tol	1		
Vector Points	25	Scale ABS	1		
Max Iterations	1000	Sensitivity	Off	=	
Max Sims		Multiple Start	Off	#Restart	0

Choose optimization parameters

0<=Reference wolf growth rate<=1
0<=Reference elk per wolf<=1
0<=Relative initial elk<=2
0<=Relative initial wolves<=2
0<=Elk fractional growth rate alpha<=1
0<=Wolf mortality rate<=1

Model value of constant --

The Optimization Control File as text

```
:OPTIMIZER=Powell
:SENSITIVITY=Off
:MULTIPLE_START=Off
```

<bla bla bla – algorithm control settings>

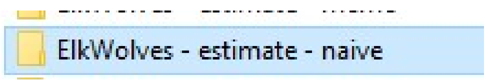
List of parameters to optimize:

```
0<=Reference wolf growth rate<=1
0<=Reference elk per wolf<=1
0<=Relative initial elk<=2
…
```

Min <= Variable Name = Initial Guess <= Max

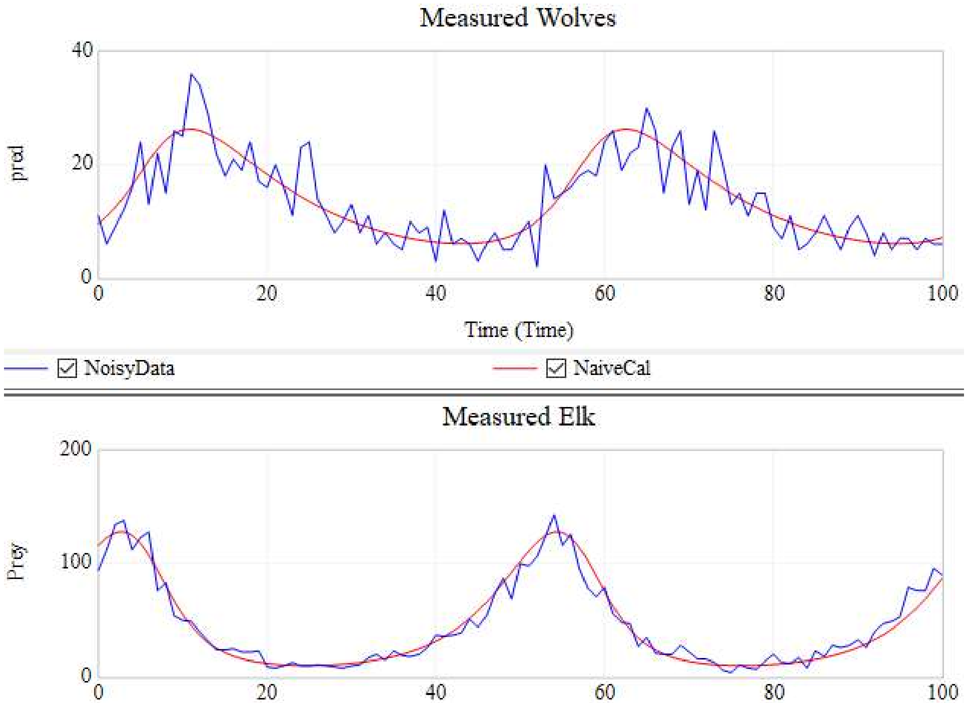
Subscript ranges are OK. Initial Guess is often omitted.

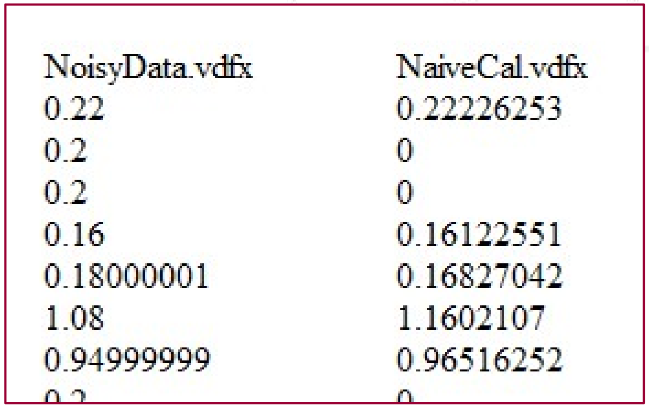
Optimize
Sample results are in


ElkWolves - estimate - naive

- The model now fits the data (hopefully)
- It fits in the future, after the short data runs out (we probably made this too easy)

- Verify: use the Runs Compare tool to see if the parameters match the synthetic data model





NoisyData.vdfx	NaiveCal.vdfx
0.22	0.22226253
0.2	0
0.2	0
0.16	0.16122551
0.18000001	0.16827042
1.08	1.1602107
0.94999999	0.96516252
0.2	0

Variations

- **Try again with …**
  - an even shorter input data series
  - more noise in the generator
  - the 2nd order model data
  - a longer forecast horizon