*The 86th Annual Meeting of the TRB*

# Application of Modern Regression Techniques and Artificial Neural Networks to Pavement Prediction Modeling

**Dr. Ying-Haur Lee, Tamkang Univ.**
**Mr. Yao-Bin Liu, National Central Univ.**
**Dr. Hsiang-Wei Ker, Chihlee Inst. of Tech.**
**Taiwan, January 21~25, 2007**

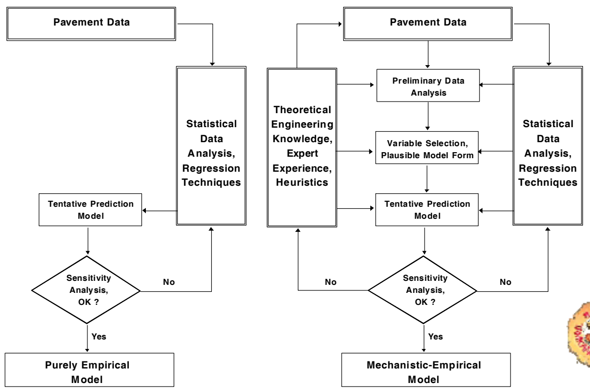


## INTRODUCTION

- Prediction Models: (pavement analysis, design, rehabilitation, PMS)
- Model Development Using Purely Empirical and Mechanistic-Empirical Approaches
- Systematic Statistical and Engineering Approach (Lee, 1993)



## Model Development Using Purely Empirical or Mechanistic-Empirical Concept (Lee, 1993)






## Previous Work on Pavement Prediction Modeling

- **Application of Modern Regression Techniques**
  - Using conventional "parametric" linear and nonlinear & several "robust" and "nonparametric" regression techniques (Lee, 1993; etc.)
  - Developed pavement performance and structural response prediction models
- **Application of Artificial Neural Network (ANN) Techniques**
  - Pavement structural evaluation for simulated data:
    - Often use original input parameters to generate the training and testing data.
    - Some parameters were fixed to certain prescribed values to reduce the database size. Result in limiting the inference space of the resulting model.
    - Nevertheless, some literature also illustrated the advantages of using the principles of dimensional analysis when generating the data.
  - Some built-in functions including learning rate and momentum term which form key neural network algorithm were not adequately investigated (Attoh-Okine, 1994;1999)
  - Adding many hidden layers gets the network to learn faster and the mean square error becomes a little smaller, but the generalization ability of the network reduces. (Sorsa et al., 1991)

## Previous Work on Pavement Prediction Modeling (continue …)

- Ripley (1993) discussed many statistical aspects of neural networks and tested it with several benchmark examples against traditional and modern regression techniques and concluded that **in one sense neural networks are little more than non-linear regression and allied optimization methods**.
- "That two-layer networks can approximate arbitrary continuous functions does not change the validity of more direct approximations such as **statistical smoothers, which certainly 'learn' very much faster**" (Ripley, 1993).
- **Statistical and subject-related knowledge** can be used to guide modeling in most real-world problems and so enable much more convincing generalization and explanation, in ways which can never be done by **'black-box' learning systems** (Ripley, 1993).



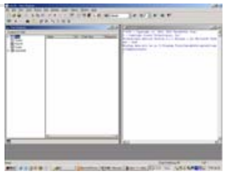
## OBJECTIVES

- To illustrate the benefits of incorporating the principles of dimensional analysis, subject-related knowledge, and statistical knowledge into pavement prediction modeling process
  - Using local regression & ANN techniques
- Case Studies:
  - To improve the prediction accuracy of simulated pavement deflections (using factorial 2-D and 3-D finite element runs and BISAR runs for different pavement systems)



## Modern Regression & ANN Techniques

- Projection Pursuit Regression
  - Revised Two-Step Modeling Approach Using PPR
- Locally-Weighted Regression (loess)
  - Concept of loess k-d tree algorithm
- Statistical Software Used
  - S-PLUS 6.1
  - LOCFIT Program
- Artificial Neural Networks
  - QNET2000 Program



S-PLUS 6.1

QNET2000 Program



## Projection Pursuit Regression *(PPR)*

$$y = \bar{y} + \sum_{m=1}^{M_0} \beta_m \, \phi_m (a_m^T x) + \varepsilon$$

$$E\left[\phi_m (a_m^T x)\right] = 0 \, , \, E\left[\phi_m^{\,2} (a_m^T x)\right] = 1$$

***Minimizing the mean squared residuals:***

$$E\left[r^2\right] = E\left[ y - \bar{y} - \sum_{m=1}^{M_0} \beta_m \phi_m \, (a_m^T x) \right]^2$$

* capable of modeling variable interactions (Friedman and Stuetzle, 1981)

## Revised Two-Step Modeling Approach Using PPR (Lee & Darter, 1994)

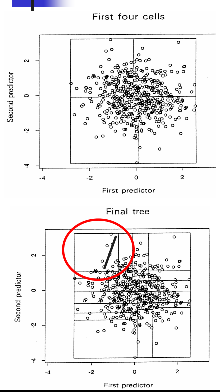
- Step 1:
  - **Use Projection Pursuit Regression (PPR)**
  - Model the multi-dimensional response surface as a sum of several smooth projected curves, graphically representable in 2-D.
- Step 2:
  - Plausible functional forms and applicable boundary conditions may then be easily identified and specified.
  - **Traditional linear, piecewise-linear, and nonlinear regressions** are then utilized to model each projected curve.
- Revised Step 2:
  - **Regression spline algorithm** was adopted here to assure smooth junctions at the change points.

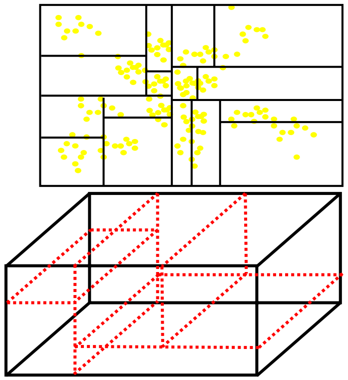
## Application of Locally-Weighted Regression (LOESS) Technique

- **An approach to regression analysis by local fitting (Cleveland & Devlin ,1988; Cleveland & Grosse, 1991)**
- **A particular data structure called k-d tree is used for partitioning space by recursively cutting cells in half by a hyperplane orthogonal to one of the coordinate axes.**
- **Use a smoothing technique for fitting a nonlinear curve to the data points locally, so that any point of the curve depends only on the observations at that point and some specified neighboring points.**
- **Provide much greater flexibility in fitting a multi-dimensional response surface as a series of many sub-divided regions with single smooth functions of all the predictors.**

10

## Illustration of loess k-d tree algorithm (Cleveland & Grosse, 1991)

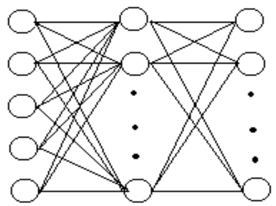


11

## Artificial Neural Networks

- A flexible way to generalize linear regression functions (but with so many parameters)
- Commonly using generalized delta rule or the steepest descent of gradient method (Back Propagation Network, BPN)
- Training & Testing Data (Over Learning Problem)

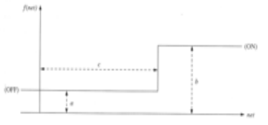



$$Min. \frac{1}{2}\sum_{p}\left\| y^{p} - c^{p} \right\|^{2}$$

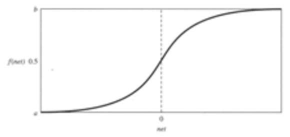
12

## Various Activation (or Transfer) Functions

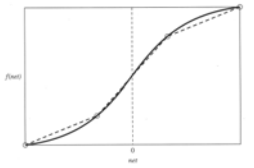
**(a) Step function**



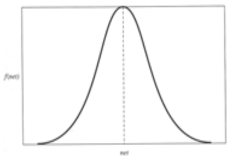
**(b) Logistic or sigmoid function**



**(c) Hyperbolic tangent function**



**(d) Radial basis function**



13

## Applications of ANN & Modern Regression Techniques

- Rigid Pavement Deflection Prediction Models
  - Case I: 2-D Infinite slab
  - Case II: 2-D Finite slab
  - Case III: 3-D Finite slab

$$R_1 = \frac{\delta_{FEM}}{\delta_{max}} = f_1\left(\frac{a}{\ell}, \frac{r}{\ell}\right)$$

$$R_2 = \frac{\delta_{FEM}}{\delta_{max}} = f_2\left(\frac{a}{\ell}, \frac{r}{\ell}, \frac{L}{\ell}, \frac{W}{\ell}\right)$$

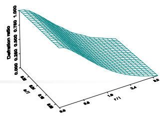
$$\frac{1}{R} = \frac{\delta_{Westergaard}}{\delta_{3D}} = f_3\left(\frac{a}{\ell}, \frac{L}{\ell}, \frac{h}{a}\right)$$

$$\ell = \sqrt[4]{\frac{Eh^3}{12(1-\mu^2)k}}$$

- Flexible Pavement Deflection Prediction Models
  - Case IV: BISAR Runs

14

## Case I: Data Preparation



- ILLI-SLAB FE Program
- Input parameters:
  - P=40 kN,
  - p=0.62 MPa,
  - E= 13.78~48.23 GPa,
  - k=13.5~175.5 MN/m$^3$,
  - h= 15.2~76.2 cm
  - r determined by mesh

  (N=12,329)
- Using Dimensional Analysis
  - a/ℓ : 0.05~0.4 (step 0.01)
  - L/ℓ=W/ℓ=8
  - r/ℓ : 0~3.2 determined by mesh generation

  (N=494)

15

## Case I: Comparing ANN Models

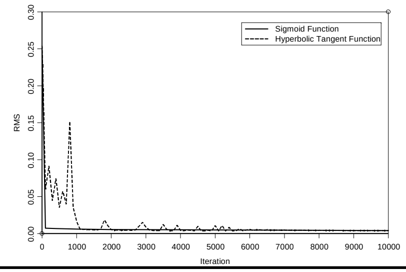
| ANN Type | NET1 | NET2 |
|---|---|---|
| Outputs | R | R |
| Inputs | E, k, h, r | *a/ℓ, r/ℓ* |
| Data Points | Training: 11,329<br>Monitoring: 1,000 | Training: 394<br>Monitoring: 100 |
| Hidden Layer (s) | 2 | 1 |
| Neurons in Each Hidden Layer | 12-12 | 6 |
| Learning Cycle | 30,000 | 10,000 |
| Learning Rate | 0.5 | 0.1 |
| Modeling Time | 6 hrs 43 min. | 42 min. |
| RMS | Training: 0.00290<br>Monitoring: 0.00420 | Training: 0.00377<br>Monitoring: 0.00360 |
| Coefficient of Determination, $R^2$ | 0.999 | 0.9999 |

**Note: Benefit of Using Dimensional Analysis (smaller & faster)**

16

## Case I: Convergence Characteristics

| Function | Logistic or sigmoid | Hyperbolic tangent | Radial basis / Step |
|---|---|---|---|
| Training RMS | 0.00416 | 0.00405 | Cannot converge |
| Monitoring RMS | 0.00384 | 0.00411 | NA |
| R-Squared | 0.9999 | 0.9999 | NA |
| Time | 35" | 60" | NA |

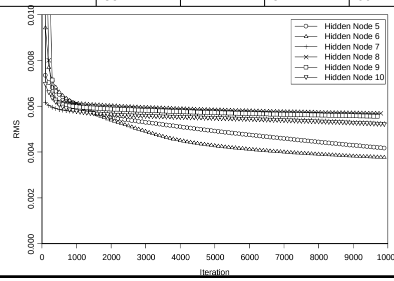



**Note: Convergence characteristics of various transfer functions**



## Case I: Convergence Characteristics
(continue …)

| Neurons in hidden layer | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|
| Training RMS | 0.00416 | 0.00377 | 0.00524 | 0.00569 | 0.00554 | 0.00520 |
| Monitoring RMS | 0.00384 | 0.00360 | 0.00492 | 0.00529 | 0.00520 | 0.00490 |
| R-Squared | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |
| Time | 35" | 42" | 52" | 60" | 67" | 82" |



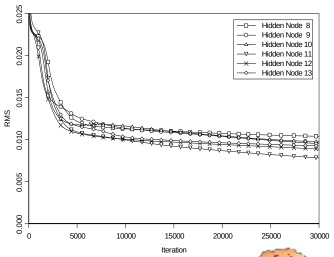

**Note:**
**Increase the # of neurons does NOT necessarily improve the fit.**



## Case II: Data Preparation

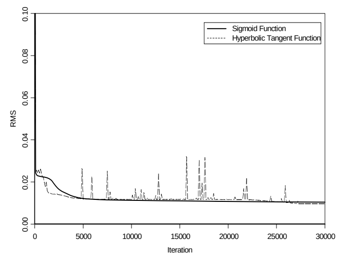
- ILLI-SLAB FE Program
- Using Dimensional Analysis
  - a/ℓ : 0.05~0.4
  - L/ℓ : 2~7 (Step 1)
  - W/ℓ : 2~7 (Step 1)
  - r/ℓ : 0~3.2 determined by mesh generation

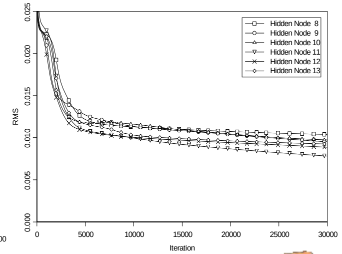
  (N=2,227)





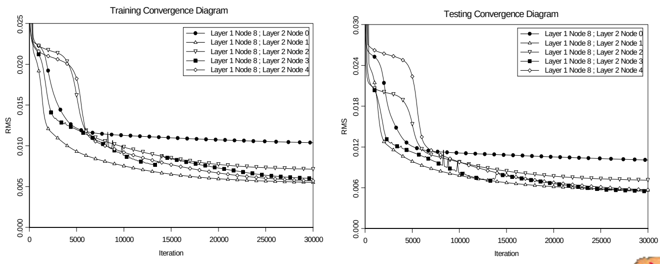

## Case II: Convergence Characteristics




**Note: due to (a) various transfer function; (b) different # of neurons (with only 1 hidden layer)**

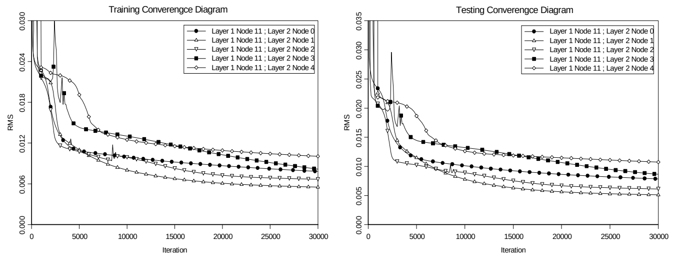
## Case II: Convergence Characteristics
(Continue …)




**Note: Two hidden layer (with 8 neurons in layer 1) converges ok!**



## Case II: Convergence Characteristics
(Continue …)

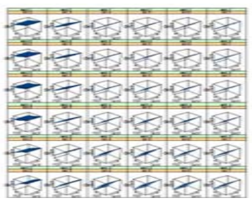


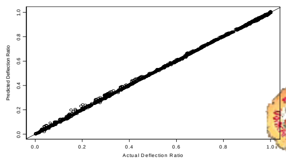

**Note: Using higher # of hidden layers and neurons sometimes lead to even worse fit, i.e., indication of over training to be avoided.**



## Case II: Loess Model

- Smoothing Parameter:
  - span=0.1
  - cell=0.01
- Regression Statistics:
  - N = 2,227
  - equivalent number of parameters = 31.9
  - SEE = 0.006376
  - R-squared = 1
- $R_2 = \frac{\delta_{FEM}}{\delta_{\max}} = f_2\left(\frac{a}{\ell}, \frac{r}{\ell}, \frac{L}{\ell}, \frac{W}{\ell}\right)$



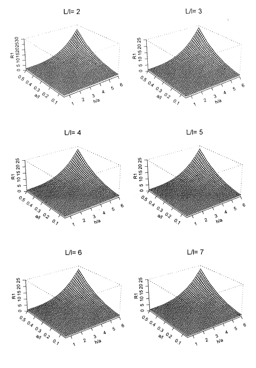




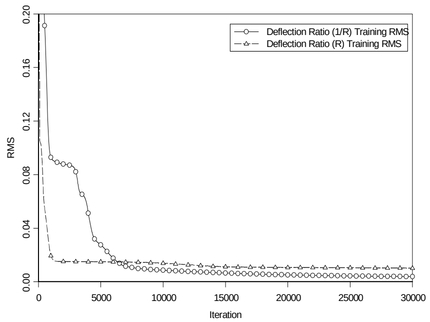
## Case III: Data Preparation

- ABAQUS 3-D FE Program
- Using Dimensional Analysis
  - a/ℓ：0.05, 0.1~0.5 (step 0.1)
  - L/ℓ=W/ℓ：2~8 (step 1)
  - h/a：0.5~6 (step 0.5)
  - Maximum deflection (r=0)

(N=504)



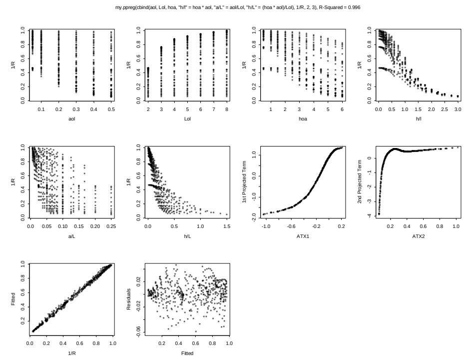
## Case III: Comparing ANN Models




**Note: Incorporating Subject Related Knowledge (1/R is better)**



## Case III: Proposed PPR Model

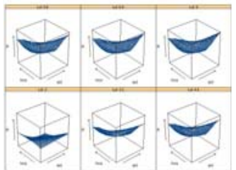


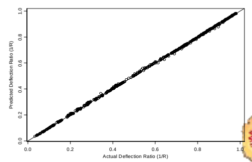
**Note: With regression spline ($R^2$=0.9942, SEE=0.02241)**



## Case III: Loess Model

- Smoothing Parameter:
  - span=0.1
  - cell=0.1
- Regression Statistics:
  - N = 504
  - equivalent number of parameters = 56.6
  - SEE = 0.004784
  - R-squared = 1
- $\frac{1}{\mathrm{R}} = \frac{\delta_{Westergaard}}{\delta_{3D}} = f_3\left(\frac{a}{\ell}, \frac{L}{\ell}, \frac{h}{a}\right)$

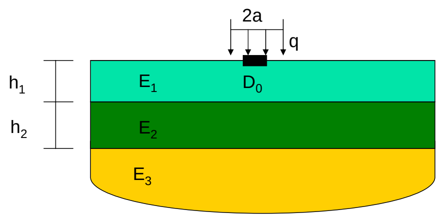






## Case IV: Data Preparation

- Factorial BISAR Runs (Flexible Pavement Deflection)
  - $a/h_2$: 0.2, 0.4, 0.8, 1.2, 1.8, 2.4
  - $h_1/h_2$: 0.5, 1.0, 1.5, 2.0, 4.0, 5.0
  - $E_2/E_3, E_1/E_2$: 0.5, 1.0, 2.0, 5.0, 10, 30, 50, 90, 140, 170
  - Training Data = 3,600, Testing Data = 1,728



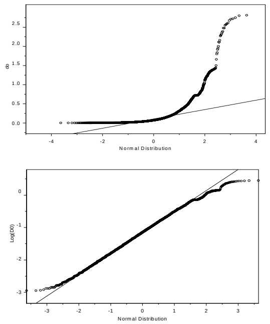
## Case IV: Comparing ANN Models

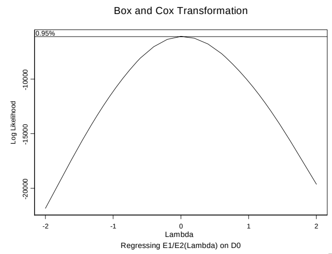
| ANN Type | NET1 | NET2 | NET3 |
|---|---|---|---|
| Outputs | $D_0$ | $Log(D_0)$ | $Log(D_0)$ |
| Inputs | $E_1/E_2$, $E_2/E_3$, $h_1/h_2$, $a/h_2$ | $E_1/E_2$, $E_2/E_3$, $h_1/h_2$, $a/h_2$ | $\log(E_1/E_2)$, $\log(E_2/E_3)$, $h_1/h_2$, $a/h_2$ |
| Hidden Layer(s) | 3 | 3 | 2 |
| Neurons in Each Hidden Layer | 20-10-5 | 15-10-5 | 12-6 |
| Learning Cycle | Cannot converge | 200,000 | 27,000 |
| Modeling Time | > 24 hrs | 10 hrs | 26 min |
| RMS | --- | Training: 0.0048<br>Monitoring: 0.0045 | Training: 0.0040<br>Monitoring: 0.0039 |

**Note: Benefit of Incorporating Statistical Knowledge (Power Transformation)**

29

## Benefit of Incorporating Statistical Knowledge



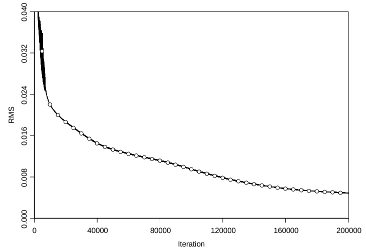




**Note: Normality test (Q-Q plot) & Box-Cox Power Transformation**

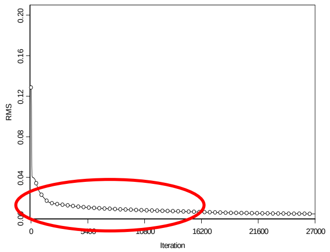
30

## Comparing Convergence Characteristics

NET2



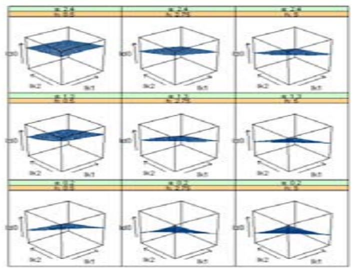

NET3




31

## Case IV: Loess Model

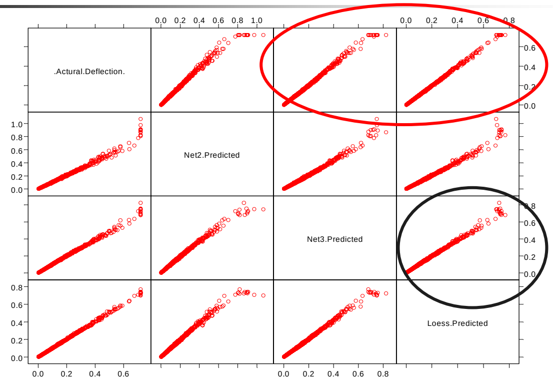
- Smoothing Parameter:
  - span=0.1
  - cell=0.1
- Regression Statistics:
  - N = 3,600
  - equivalent number of parameters = 31.9
  - SEE = 0.02792
  - R-squared = 1



- $$\log(D_0) = f_4\left(\log(\frac{E_1}{E_2}),\quad \log(\frac{E_2}{E_3}),\quad \frac{h_1}{h_2},\quad \frac{a}{h_2}\right)$$

32

## Model Comparison (Testing Data)




33

## Concluding Remarks

- Illustrated the benefits of incorporating
  - the principles of dimensional analysis,
  - subject-related knowledge, and
  - statistical knowledge

  into pavement prediction modeling process
- Proved to have higher accuracy
- Required smaller data and less network training time
- Increasing the complexity of ANN models does NOT necessarily improve the fit

34

## Concluding Remarks (Continue ...)

- Using higher # of neurons and hidden layers sometimes lead to even worse fit (indication of over training to be avoided)
- Reasonable good predictions can be achieved using both ANN and modern regression techniques
- Statistical and subject-related knowledge can be used to guide modeling and so enable much more convincing generalization and explanation, in ways which can never be done by "black-box" learning systems (Ripley, 1993)

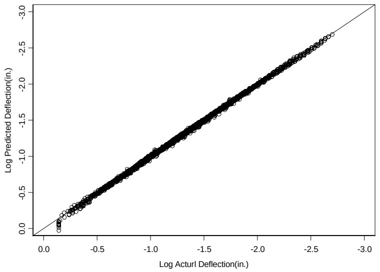
35

Questions?

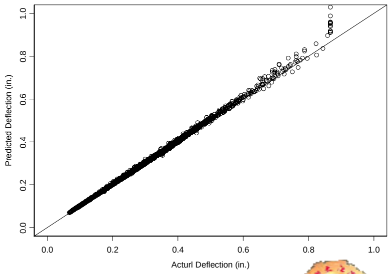
THANKS FOR YOUR ATTENTION

36

## NET2 Goodness of Fit (NET2)



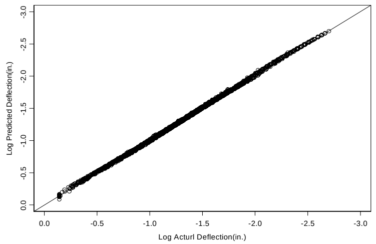

Predicting Log ($D_0$)



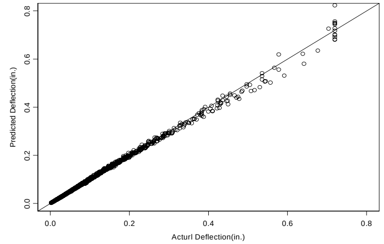

Predicting $D_0$

37

## NET3 Goodness of Fit



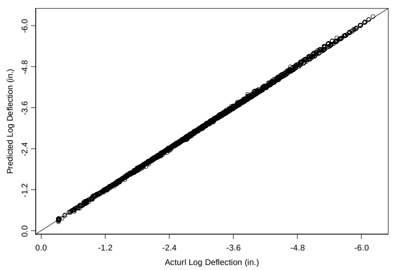

Predicting Log ($D_0$)



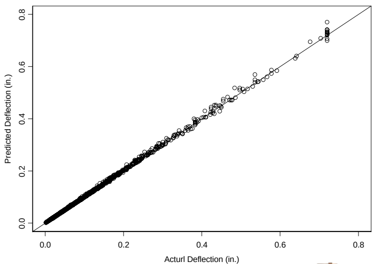

Predicting $D_0$

38

## Loess Model: Goodness of Fit




Predicting Log ($D_0$)




Predicting $D_0$

39