A Quantitative Analysis of Factors Affecting Broken Rails

Darwin H. Schafer II & Christopher P.L. Barkan

May 9th, 2008

The William W. Hay Railroad Engineering Seminar Series

University of Illinois at Urbana-Champaign

CN BNSF RAILWAY ILLINOIS RAILROAD ENGINEERING PROGRAM

Introduction to Broken Rails

- Broken rails are caused by inherent flaws (internal defects) in the rail and/or by fatigue on the surface of the rail.
  - A **broken rail derailment** is the occurrence of a train accident due to the rail failure
  - A **service failure** is a detected broken rail event without the occurrence of a train derailment.
    - Usually detected by the signal system

Probability of Broken Rail Events

Broken rails result from the growth of internal or surface defects:

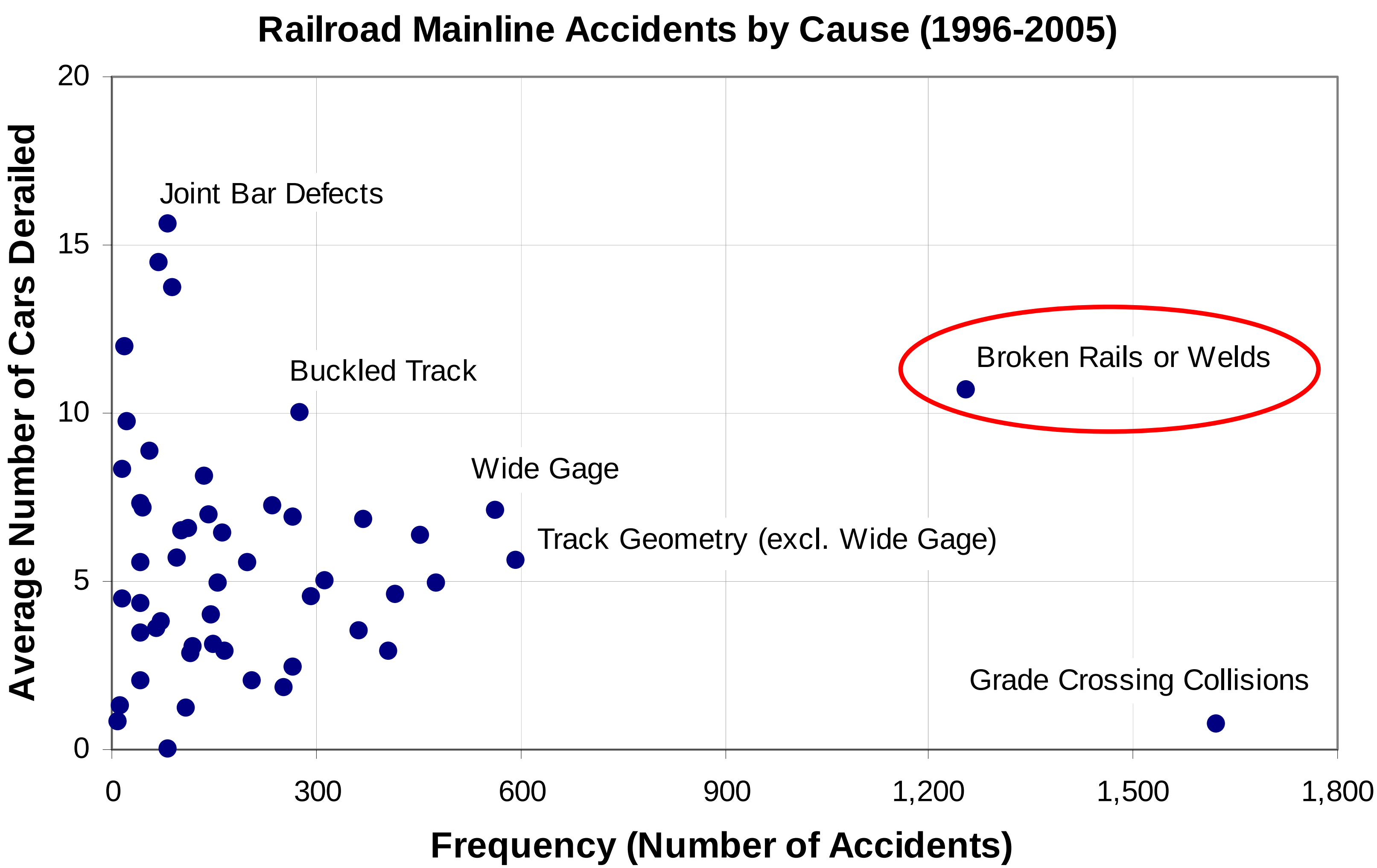
- Internal rail defects typically start as microscopic imperfections during the manufacturing of the rail.
  - Defects grow slowly due to cyclic loading and unloading of the rail from traffic and thermal stresses
  - Under the “right” conditions they may grow rapidly leading to a service failure or broken rail derailment
- The growth of defects is linked to a number of variables:
  - Wheel loads, types of rail, traffic levels, maintenance activities, changes in track modulus, etc.

Consequences of Broken Rail Events

Detection of broken rails is vital:

- Broken rails are the leading causes of major train derailments in U.S. and accident-caused hazardous materials releases
- From 2003 through 2006 broken rails accounted for:
  - 335 mainline derailments for Class I railroads in the U.S.
  - Over $176 million of equipment & track damage
    - Approximately $525,000 per incident
  - 14 hazardous material release accidents
- Service failures also cause delay and track repair cost
  - This is especially true on high density lines

Importance of Broken Rail Prevention



Presentation Outline

- Economic Impact
  - Cost associated with broken rail derailments and service failures
  - Cost associated with train delay
  - Cost of typical preventive measures
- Prediction of broken rails
  - Description of broken rail data
  - Development of new statistical and neural network prediction models
  - Presentation of practical model for broken rail prediction
- Conclusions & future work

Presentation Outline

- Economic Impact
  - Cost associated with broken rail derailments and service failures
  - Cost associated with train delay
  - Cost of typical preventive measures
- Prediction of broken rails
  - Description of broken rail data
  - Development of new statistical and neural network prediction models
  - Presentation of practical model for broken rail prediction
- Conclusions & future work

Economic Impact of Broken Rails

Costs of rail failures include (Cannon et al. 2003):

- Cost of broken rail derailments
  - Track & equipment damage
  - Clean-up costs, lading damage, loss of business, etc.
- Remedial treatments
  - Replacement or repair of rails and welds
- Train delay
  - Based on time out-of-service and line density
- Track inspection
  - Ultrasonic & geometric testing
- Preventive methods
  - Rail grinding, rail replacement, track surfacing

Cost of Broken Rail Derailments

Reportable accidents to the FRA:

- Accident cost must be greater than $7,700 (for 2006)
- Equipment damage reported to the FRA
  - Repair or replacement of on-track rail equipment.
    - Cars, locomotives, and maintenance equipment
  - Includes cost labor and materials needed.
- Track damage reported to the FRA
  - Repair or replacement of any track, signals, or track structures.
  - Takes into account material and labor costs associated with clearing the right-of-way and repairing the roadbed .
  - Includes contracted third-party labor fees.

Cost of Broken Rail Derailments

FRA reportable broken rail accidents:

- Class I US freight railroad mainline accidents
  - 335 broken rail derailments from 2003 - 2006
  - Total cost of $176 million
    - Average cost of broken rail derailment is $525,400

FRA Code	Cause Description	Frequency	Total Cost	Cost Per Incident
T201	Bolt hole crack or break	14	$12,854,596	$918,185
T202	Broken Base	26	5,804,332	223,244
T203	Broken Weld (plant)	3	1,026,794	342,265
T204	Broken Weld (field)	25	17,338,957	693,558
T207	Detail fracture from shelling or head check	82	52,792,131	643,806
T210	Head and web separation (outside joint bar)	23	5,380,144	233,919
T211	Head and web separation (within joint bar)	7	2,042,042	291,720
T212	Horizontal split head	5	1,967,657	393,531
T213	Joint bar broken (compromise)	6	3,111,204	518,534
T214	Joint bar broken (insulated)	9	8,152,304	905,812
T215	Joint bar broken (noninsulated)	15	14,225,856	948,390
T219	Rail defect with joint bar repair	1	664,622	664,622
T220	Transverse/compound fissure	90	42,315,036	470,167
T221	Vertical split head	29	8,333,190	287,351
		335	**$176,008,865**	**$525,400**

Cost of Broken Rail Derailments

Costs not reported to the FRA:

- Train delay
  - Depends on severity of accident, access to site, if hazmat is involved or if near a metropolitan area
  - One railroad estimates that a “moderate to large” broken rail derailment has an average of 24 hours of track outage.
    - Typically broken rail derailments pile cars up in the location at the point of derailment and only damage track in that immediate area.
  - Cost can be approximated using developed train delay calculator
- Lading damage
  - Not reported to FRA, future research needed.
- Loss of business or customer support
  - Very difficult to quantify, future research needed.

Cost of Broken Rail Service Failures

Costs associated with broken rail service failures:

- Material and labor costs:
  - Railroad estimates range from $750 to $1,500 for material, labor, and mobilization costs to repair a service failure.
    - The cost of mobilization varies significantly based on availability of material and access to the site.
- Train delay costs:
  - Length of delay depends on accessibility to the site, repair crew availability, and if proper materials are readily available.
  - One railroad estimates that the average track-outage for a broken rail service failure is 3-4 hours from the time of notification.
  - Delay cost can be estimated using train delay calculator.

Cost of Train Delay

Breakdown of estimated train delay cost:

- Cost based on four categories (calculated from industry operating averages):

  1) Car cost: **$20.67 per train-hour**
     - Cost of delaying railroad owned cars that cannot be used.
     - Depends on % or railroad owned cars in train consist.

  2) Locomotive Cost: **$64.54 per train hour**
     - Cost from loss of the use of locomotives delayed.

  3) Fuel Cost: **$18.24 per train-hour**
     - Fuel used while train is waiting in idle.
     - Depends on locomotive type.

  4) Crew Cost: **$110.08 per train-hour**
     - Train crew labor cost during a delay (including fringe benefits).

Cost of Train Delay

Total train delay cost:

- Calculated total cost per train-hour of delay is **$213.52**.
  - Value calculated based on industry averages for Class I railroads.
  - Railroads in the past have used train delay cost values ranging from $200 to $300 per train-hour.
    - Varies based on type of train, geographic location, etc.
- Additional train delay cost sources that were not considered:
  - Crew replacement due to federal hours-of-service regulations.
  - Fuel consumed for acceleration after train stoppage.
  - Delay cost of privately owned cars.
  - Cost of delay for multiple trains.

Train Delay Cost Calculator

Train delay due to multiple trains:

- Will vary based on the density of the line, the number of mainline tracks, and the length of the delay.
- Assume that trains will arrive in constant time intervals
  - This assumption is less valid for low density lines.
  - Interval between train arrivals can be calculated based on the number of Annual Million Gross Tons (MGTs)
    - Average train for Class I freight RR's is 6,312 gross tons.

$$n = \text{Number of trains per year} = \frac{\text{Annual MGTs}}{\text{tons per train (millions)}} = \frac{ANMGT}{0.006312}$$

$$t = \text{hours per train arrival} = \frac{\text{hours per year}}{\text{trains per year}} = \frac{8{,}766}{n} = \frac{55.33}{ANMGT}$$

Train Delay Cost Calculator

Train delay cost formula:

$$C = Tx + \sum_{n=1}^{m} (T - nt)x$$

where,

C = total train delay cost for multiple trains

T = total delay time for service interruption

x = cost of delay per train-hour ($213.52)

m = number of following trains delayed = T / t (rounded to the nearest integer)

t = hours per train arrival = 55.33 / ANMGT

- Train delay scenarios:
  1) Single track service failure or derailment (100% delay)
  2) Multiple track broken rail derailment (100% delay)
  3) Multiple track service failure (50% or less of delay)
     – Adjustment factor needed based on the density of the line.

Cost of Preventive Measures

Typical practices for preventing broken rails:

- Rail inspection
  - Ultrasonic inspection to search for internal defects in rail.
  - Inspection to detect geometric defects of the track.
- Rail grinding
  - Grind rail and remove surface defects – extends life of rail.
- Rail surfacing
  - Maintains stable and properly aligned track.
- Rail renewal and replacement
  - Occurs due to both wear on the rail and crack growth.
  - Replaced at local level or by capital rail replacement projects.

Cost of Preventive Measures

Broken Rail Prevention Technique	Annual Cost per Track Mile ($)
Ultrasonic and Geometric Track Inspection	900
Rail Grinding	1,900
Rail Surfacing	700
Rail Renewal and Replacement	2,500
TOTAL INDIRECT COST	**6,000**

- Cost estimated to be $6,000 per track mile maintained.
  - Based on industry reported estimates and averages.
- The cost of preventive measure are considered to be indirect costs of broken rails:
  - These practices have other benefits unrelated to broken rail prevention.
    - Rail grinding and surfacing extends overall rail life
    - Rail replacement is completed for many reasons

Summary of Economic Impact

- Based on this analysis estimated costs for broken rails:
  - Broken rail derailment: $550,000 and higher per incident
  - Service failure: $2,500 and higher per incident
  - Indirect cost of preventive measures: $6,000 per track mile
- Estimated annual costs for all Class I railroads:

Broken Rail Associated Cost	Cost ($)
Derailment Damage	44,002,216
Derailment Delay Cost	3,006,793
Service Failure Repair Cost	19,141,747
Service Failure Delay Cost	16,768,170
TOTAL DIRECT COST	**82,918,926**
Track Inspection	128,185,200
Rail Grinding	270,613,200
Rail Surfacing	99,699,600
Rail Renewal and Replacement	356,070,000
TOTAL INDIRECT COST	**854,568,000**

Presentation Outline

- Economic Impact
  - Cost associated with broken rail derailments and service failures
  - Cost associated with train delay
  - Cost of typical preventive measures
- Prediction of broken rails
  - Description of broken rail data
  - Development of new statistical and neural network prediction models
  - Presentation of practical model for broken rail prediction
- Conclusions & future work

Problem Description

Objective: Develop a model to identify locations in the rail network that have a high likelihood of experiencing a broken rail event.

Classification Problem:

- Failure vs. non-failure – if a broken rail event occurred
- Input variables:
  - Rail characteristics (Rail age, type, weight)
  - Track characteristics (Curvature, superelevation, gradient)
  - Traffic levels (Annual MGTs, speed, car loads, wheel passes)
  - Infrastructure (Bridges, turnouts, culverts, grade crossings)
  - Rail testing results (Geometric and ultrasonic inspections)

Data Available for Study

Service failure data used in analysis:

- 25,370 unique track sections from the BNSF Railway
  - Each track section has length 0.01 miles (~53 feet).
  - 12,685 sections experienced a rail failure in four year period.
    - 2003 through 2006
  - 12,685 randomly selected sections without a broken rail event.
- Rail network data provided for each track segment
  - 28 input variables for the model:
    - Rail characteristics, track characteristics, traffic levels, infrastructure features, and rail defect testing results

Previously Developed Statistical Model

Previous statistical model developed by Dick (2001):

- Model was formulated using broken rail data for two-year time period on the BNSF network.
  - May 1998 to May 2000 service failure data
- Logistic regression used to predict broken rail locations.
  - Dick's (2001) model stated a 87.4% level of accuracy
- Model was tested for our current service failure data
  - Found to be only 54.8% accurate for current data
  - Further examination of the data used in Dick's (2001) study revealed that the data was missing service failure locations

Updated Statistical Classification Model

Logistic regression (Logit) model development:

- Logistic regression technique used again to develop classification model with updated service failure data.
  - Logit is a discrete-choice, statistical model that constructs an equation based on the input variables to predict failure.
    - Only variables significant to the model are incorporated.
    - Output of the model is a value between 0 and 1 indicating the probability of failure.
  - Different variable selection techniques:
    - Forward selection
    - Backward selection
    - Step-wise selection
  - Tested model against “unseen” data (validation sample)

Logit Modeling Techniques

Different modeling techniques were used:

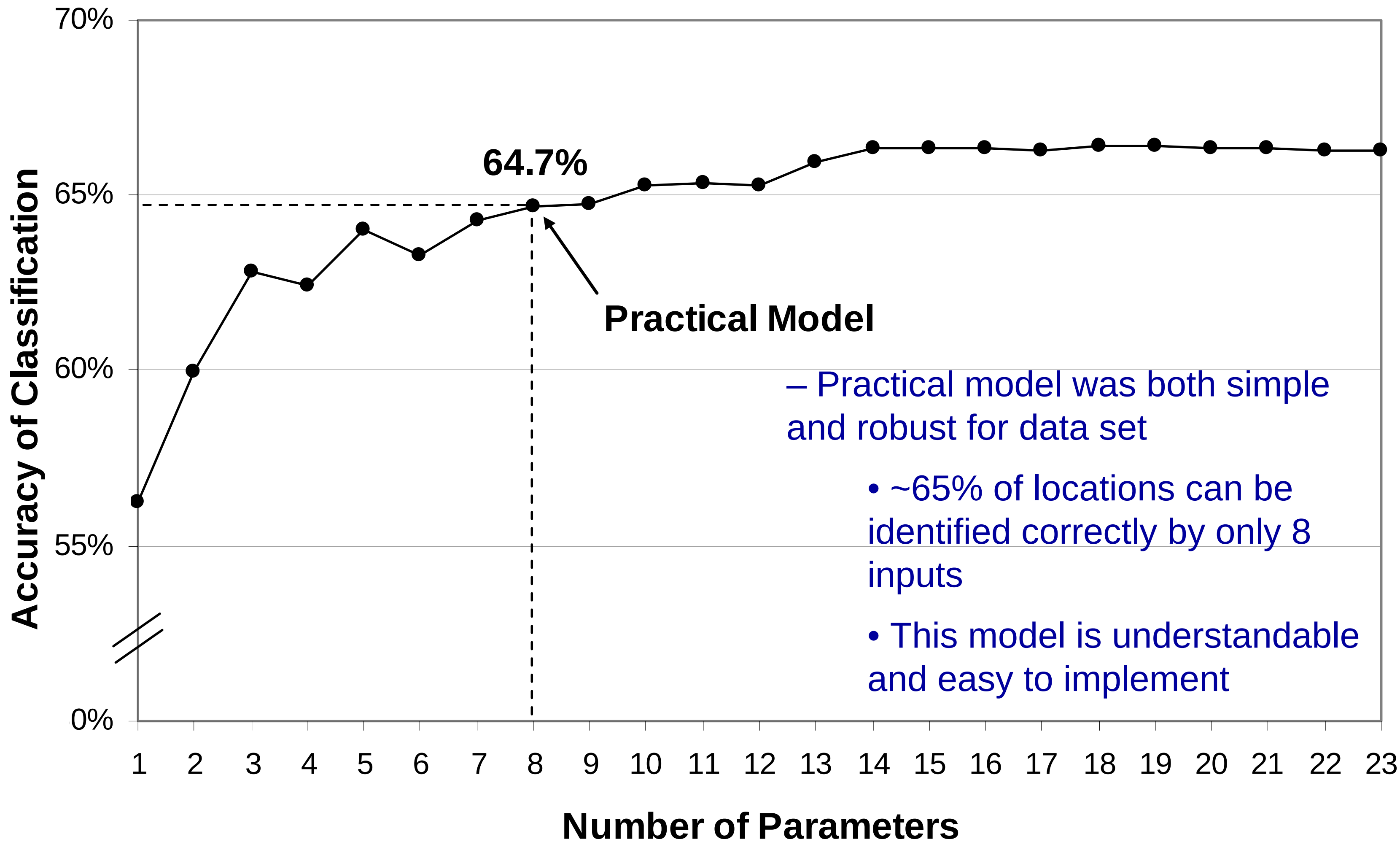
- Transformation of input parameters:
  - Changing applicable inputs from continuous to discrete variables
    - Rail weight transformed into 13 distinct rail weights
  - Binary inputs already act as discrete variables
    - Rail type (bolted or welded), occurrence of a defect, bridge present
- Multiple term interaction:
  - Model allows for variables to have interdependency
    - For example, rail age and degree of curve may have a combined effect on service failure probability
    - Computational power and resources only allow for up to two-term interaction models using logistic regression

Statistical Model Results

Logistic Regression Model	Number of Parameters	Accuracy of Initial	Accuracy for Testing Sample
Simple Logit	23	66.3%	66.3%
Simple Logit w/ Transformation	34	68.5%	67.7%
Two Term Interaction Logit	145	71.1%	57.1%
Two Term Interaction Logit w/ Transformation	336	72.3%	67.3%
Practical Logit Model	8	64.7%	64.1%

- Results:
  - Logit models doing a fairly good job but required a lot of parameters
  - Two term interaction models also appeared to be doing well, but were “over-fitting” the data
    - Not accurate for “unseen” data

"Practical" Model



– Practical model was both simple and robust for data set

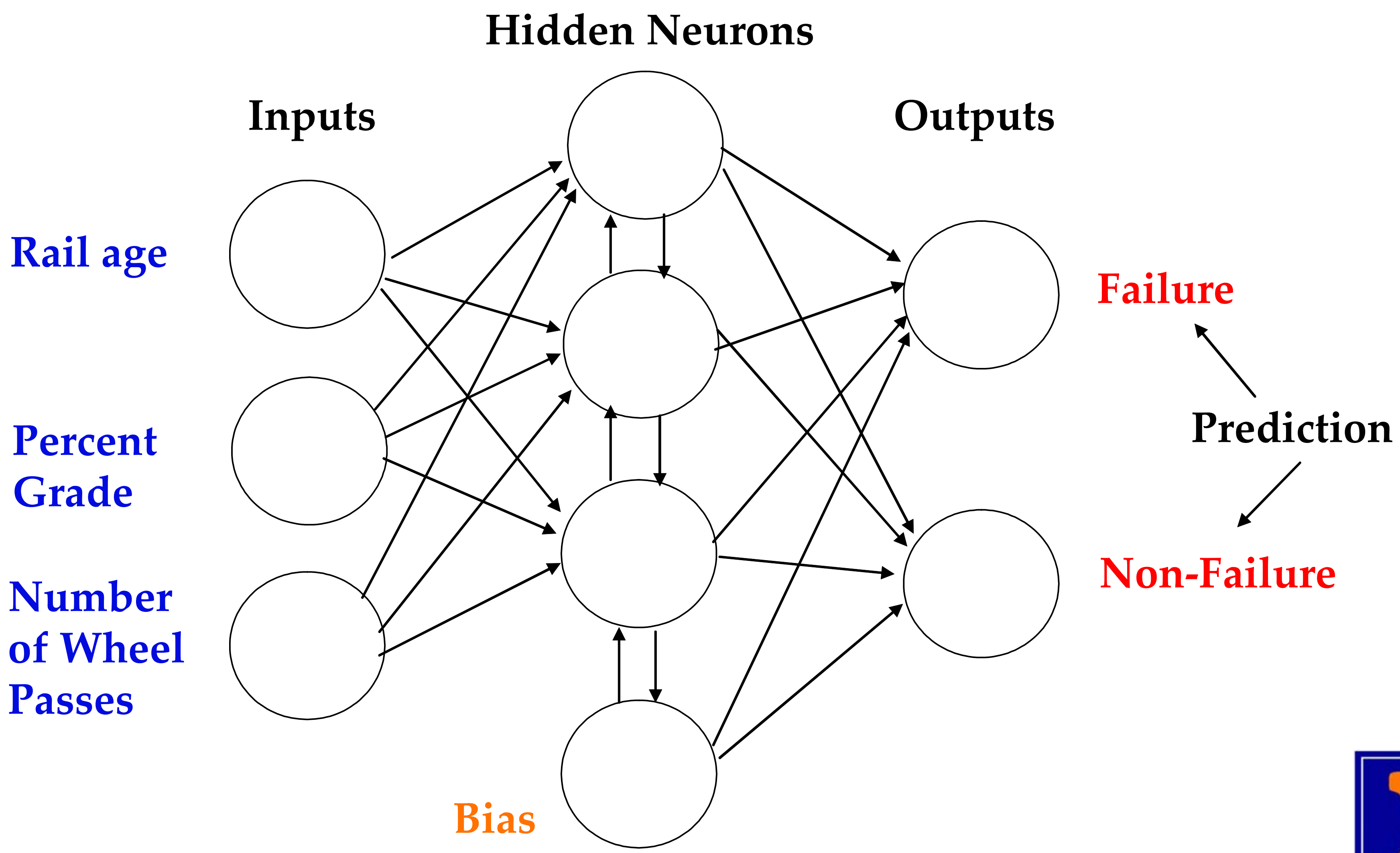
- ~65% of locations can be identified correctly by only 8 inputs
- This model is understandable and easy to implement

Artificial Neural Network Model

Artificial Neural Networks (ANN) for classification:

- Alternative to traditional logistic regression analysis:
- Advantages of Neural Networks:
  - Can detect complex non-linear relationships between input and output variables.
  - Can detect all possible interactions between predictor variables.
- Disadvantages of Neural Networks:
  - Greater computation time for large networks
  - No output value for probability – only classification
  - “Black Box” model
    - Limited ability to explain what is happening in the model and the possible causal relationships.

Artificial Neural Network



Neural Network Classification Model

Simple ANN Model:

- Built using same dataset as logistic regression model
- Learning method used was backpropagation
  - Most common learning method of ANNs
- Results:
  - 67.7% correct classification.
  - Optimal number of hidden neurons: 76
    - ANN restricts total number of neurons to not over-fit the model
  - Tested against unseen data and model was found to be robust:
    - 66.6% correct for validation sample.

Hybrid Logit/Neural Network Models

Hybrid classification models:

- Research in economic and medical applications have shown that hybrid models show a higher level of classification accuracy as compared to other techniques.
- Logistic regression analysis can be used as input information for neural network learning.
  - Two types of hybrid models:
    1) Pre-selection of input variables (Logit-ANN)
    2) Probability output as an additional input (PLogit-ANN)
- Main advantage of a hybrid model:
  - Information from Logit allows for fewer cases to be used on learning and therefore more cases are used on optimization.

Hybrid Logit/Neural Network Models

Hybrid models used for broken rail prediction:

1) Pre-selection of variables (Logit-ANN)

- 23 variables selected as significant from step-wise Logit model
  - 67.5% correct classification.
  - Optimal number of hidden neurons: 71
  - 66.6% correct for validation sample.

2) Additional input variable of probability (PLogit-ANN)

- Step-wise Logit model used to produce probability of failure.
  - 67.9% correct classification.
  - Optimal number of hidden neurons: 77
  - 66.4% correct for validation sample.

Summary of Developed Models

Model Type	Accuracy of Initial Classification Model	Accuracy for Testing Sample
Simple Logit	66.3%	66.3%
Simple Logit w/ Transformation	68.5%	67.7%
Two Term Interaction Logit	71.1%	57.1%
Two Term Interaction Logit w/ Transformation	72.3%	67.3%
Practical Logit Model	64.7%	64.1%
ANN	67.7%	66.6%
Logit-ANN	67.5%	66.3%
Plogit-ANN	67.9%	66.4%

- Results:
  - Neural network (ANN) & hybrid neural networks preformed as well as statistical models (Logit) for service failure prediction.
  - Two-term Logit models were not robust for validation data.
  - Practical Logit model is simple, understandable, and useable.

Practical Logistic Regression Model

Prospective statistical service failure model:

$$P = e^{U} / (1+e^{U})$$

where, P = probability that a service failure will occur

$$U = -0.270 - 0.0454S - 1.35R - 0.0106A + 0.00899T + 0.0232L + 1.61I + 0.823G + 1.63B$$

where,
- S = Rail weight (in pounds per yard)
- R = Rail type (1 if welded, 0 if bolted)
- A = Rail age (in years)
- T = Annual Traffic (in million gross tons)
- L = Weight of car (in tons)
- I = Presence of an ultrasonic defect (1 if present, else 0)
- G = Presence of a geometric defect (1 if present, else 0)
- B = Presence of a bridge within 200’ (1 if present, else 0)

Application of Practical Model

- The "practical" Logit model was transformed into a prospective prediction model.
  - Prospective model based on actual rate of service failures.
  - This model can be used to determine the probability of a broken rail or expected number of broken rails on a line.
- The models developed in this study can be used to assist railroads in prevention of broken rails.
  - Short-term maintenance planning:
    - Assist roadmaseters on a local scale to identify locations that have a high probability of experiencing a broken rail.
  - Long-term maintenance planning:
    - Model can be used to determine the number of expected service failures on various lines to help determine maintenance priorities.

Presentation Outline

- Economic Impact
  - Cost associated with broken rail derailments and service failures
  - Cost associated with train delay
  - Cost of typical preventive measures
- Prediction of broken rails
  - Description of broken rail data
  - Development of new statistical and neural network prediction models
  - Presentation of practical model for broken rail prediction
- Conclusions & future work

Conclusions

- Economic impact of broken rail events is significant:
  - Annual direct cost of $83 million due to broken rail service failures and derailments on US Class I railroads.
  - Annual indirect costs exceed $850 million for broken rail prevention techniques.
- Statistical, artificial neural network, and hybrid modeling techniques were used to develop prediction models.
  - ANN and hybrid models did not show a significant improvement over the stand-alone statistical models.
  - "Practical" statistical model developed is simple & useable
    - Included only the top 8 predictor variables.
    - ~ 65% accuracy achieved as well as robust for unseen data.

Future Work

- Future research may be possible to understand the economic impact of “hard-to-quantify” values:
  - Lading loss in a derailment.
  - Loss of business from accidents and train delays.
  - Costs associated with delaying privately owned cars.
- Future work may also be possible to refine and improve our broken rail prediction models.
  - May be possible my attempting to understand the remaining unexplained model variance (~30% misclassified)
  - Evaluate additional possible predictive factors for broken rails:
    - Service failure rates, defect rates, temperature fluctuation, flat wheel incidence, inspection frequency, etc.

Acknowledgments

– Current & past University of Illinois students:
  - Daniel Peltier, Yung-Cheng (Rex) Lai, Rapik Saat, Mark Dingler, Billy Lau, John Zeman, Tyler Dick, & Avery Grimes

– Industry assistance:
  - BNSF Railway: Larry Milhon, Mike Mischke, Paul Gabler, Manal Bishr, Dennis Morgart, Mark Stehly, & Ed Chapman
  - CN Railway: David Ferryman
  - Union Pacific Railroad: Dale Mellor
  - CSX Transportation: Chuck Gullakson & Michael Hill
  - TTCI: Tom Guins
  - RSI-AAR Tan Car Safety Project: Todd Treichel

– Support from CN Fellowship and BNSF Research Grant