

2000 AEIC Load Research Conference

# Validation, Editing and Expansion in a Deregulated Environment

by Joseph S. Lopes  
Applied Energy Group, Inc.



Applied Energy Group, Inc.  
2000 AEIC Load Research Conference

1

## Introduction

- **Joseph S. Lopes**
  - Math & Computer Science Degree - Manhattan
  - Applied Energy Group, Inc. Co-founder (1982)
    - Senior VP, Treasurer 1982 - Present
  - American Electric Power 1977-1980
    - Load Research Group (Bill Mekolites)
    - Preceded Jeff Laine (AEIC Chairman)
  - Stone & Webster Management Cons. (1980-82)
  - Developed PC-IDEAS LR System for NYPA 1985-90 (1st PC LR System)



Applied Energy Group, Inc.  
2000 AEIC Load Research Conference

2

## Background: Validation & Editing

- **Applications of Interval Load Data**
  - 1960's/70's: Rates and Planning
  - 1978: PURPA, Cost-of-Service
  - 1980's: DSM Evaluation
  - 1990's: Technology Assessment
  - 2000's: Load Profiling
- **Quality of Data Always an Issue**
  - Garbage in - garbage out



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

3

## Background: Validation & Editing

- **Implications of Bad Interval Data**
  - Small sample sizes expanded by multiples of 1,000 - 100,000
  - Rate and regulatory scrutiny leads to disqualification of cost recovery and lower rate of return
- **Causes of Bad and Lost Data**
  - Equipment Failure, Mis-communications, human error, weather, human error, computer failures, and bad luck



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

4

# History of Load Data

- **Prehistoric Days**

- 7,000 - 10,000 days BC (Before Competition)

- **Magnetic Tape Recorders (1960's, 1970's)**

- better than strip charts
- tapes are bulky
- tapes jam, stretch, break
- time splice mismatches
- outage timing uncertain



» Load Research Man-->



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

5

# History of Load Data

- **App. 7000 (days) B.C. (Before Competition)**

- **Load Research Man discovers Batteries!**

- Continuous time pulse
- Easier to identify outages

- **App. 3,000 B.C.**

- **LR Man discovers electronics!**

- No more mechanical errors
- Infant mortality worst problem

- **Since then...**

- optical ports, modems, wireless



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

6

## Load Data Today

- **Not all Load Research Programs survived the Dark Ages**
  - Mergers, cost-cutting, preparing for deregulation, staff “retirements”
- **Today, 2000 A.D. (After Deregulation)**
  - Digging out the old data
  - Polishing off old techniques
  - Revitalizing samples



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

7

## Why Validate and Edit?

- **High costs and customer intrusion make each data point valuable**
- **Limited budgets, manpower, priorities, other resources in “leaner” utility mean smaller samples**
- **Small margin for error, less time to fix**
- **Each data point lost could be 1% of typical small samples of 100 per class**
- **Must recover all data possible**



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

8

# Load Data Validation

- Has new technology eliminated bad data? *No, it has reduced problems and changed the types of problems*
- Objectives of Validation
  - identify data that is inaccurate, missing or not representative
    - holes in the data: missing because data lost
    - outages: accurate but not representative!
    - Spikes and troughs: Outage followed by payback

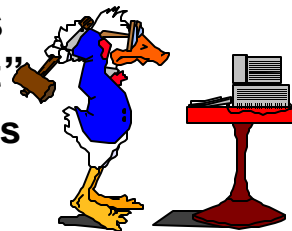


Applied Energy Group, Inc  
2000 AEIC Load Research Conference

9

# To Edit or Not to Edit?

- “Edit by Exclusion” is often the easiest “edit”
  - risk too much data loss
  - could bias the results
- Edit Patch Techniques
  - Judicious use of patching
  - Consistent criteria essential
  - Maintain natural variability of data



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

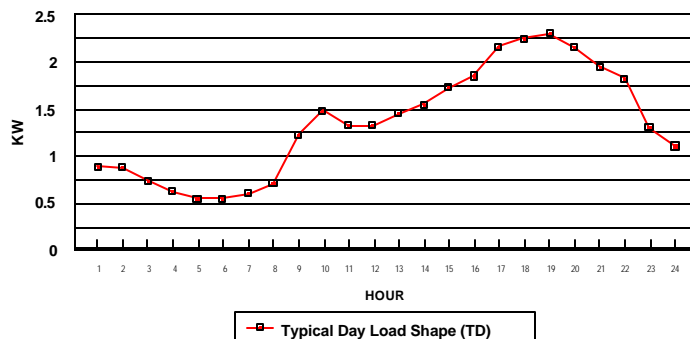
10

# Keys to Effective Validation

- Validation should:
  - Identify problem data automatically
  - Only flag a small percentage
  - Have consistent criteria
  - Minimize Errors:
    - Type I Error - Reject good data
    - Type II Error - Accept bad data
  - Type II Error is more serious

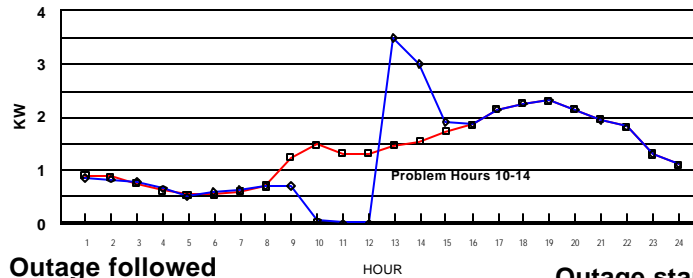
# Validation - Typical Day

RESIDENTIAL LOAD SHAPE VALIDATION ANALYSIS  
Chart A - Typical Day Load Shape (Summer Residential)



# Validation - Problem Data

RESIDENTIAL LOAD SHAPE VALIDATION ANALYSIS  
Chart B - Typical Day vs. Outage Day Load Shapes



Outage followed  
by recovery  
“payback”

Legend:  
- Typical Day Load Shape (TD)  
- Outage Day Load Shape (OD)

Outage started  
between 8-9,  
ended 12-13



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

13

# Validation Tests

- **Basic Tests and Issues**
  - Missing Data
  - Zero use intervals (unless end uses)
  - Zero use for days/weeks (data communications?)
  - Unusual Inconsistency (may be usual)
  - mismatches against billing (may not be available)
  - Best to compare against prior period

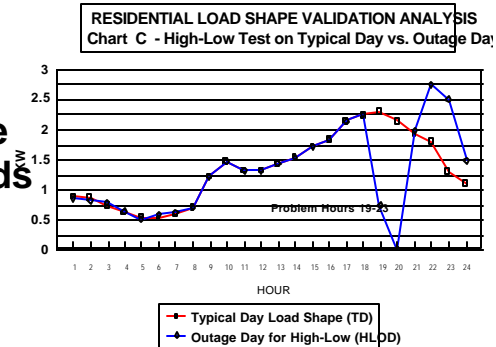


Applied Energy Group, Inc  
2000 AEIC Load Research Conference

14

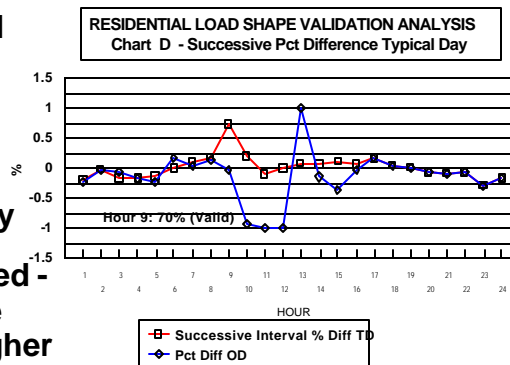
## Validation - High/Low Tests

- Typically used for Billing
- Not as effective for interval loads
- need all hours accurate, not just peak
- Cannot account properly for natural variability of data



## Validation - Successive % Difference

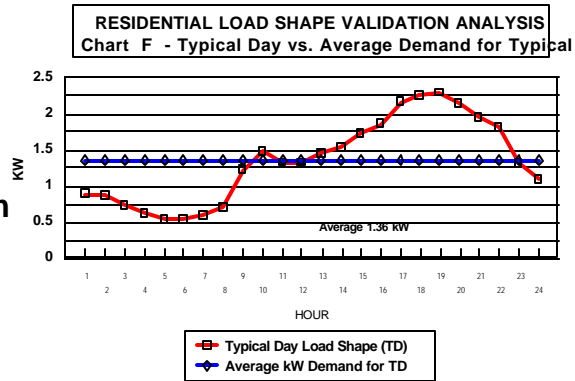
- Used for interval data tests
- Not effective for variable loads
- Cannot handle normal variability
- Inherently biased - some intervals are naturally much higher (early AM) or lower (PM) than previous



Missed Hour 14 problem on Outage Day

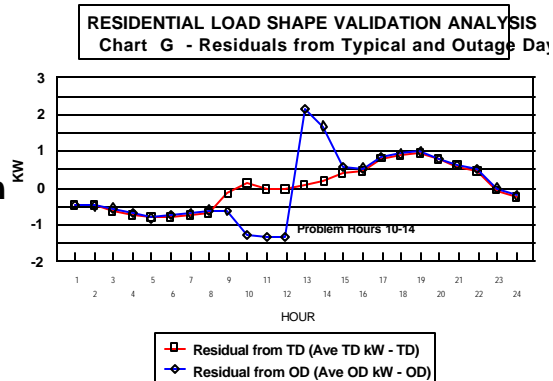
## Validation Test - Pattern Recognition

- A good concept, but how to do it
- Option 1: Create a residual from the average daily interval load



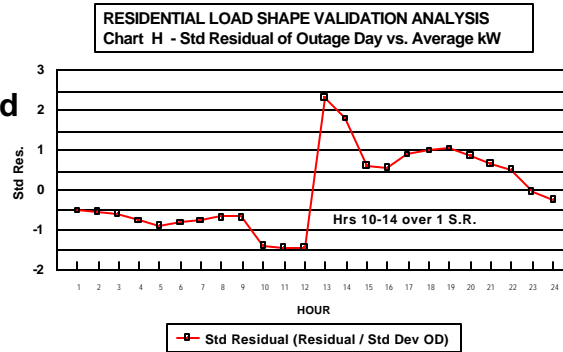
## Validation Test - Pattern Recognition

- Compare residual from typical and outage days
- Set a criterion that will consistently distinguish good from bad intervals



## Validation Test - Pattern Recognition

- **Option 2:**  
Create standardized residual by dividing residual by standard deviation
- **Identify Outage Day bad intervals**

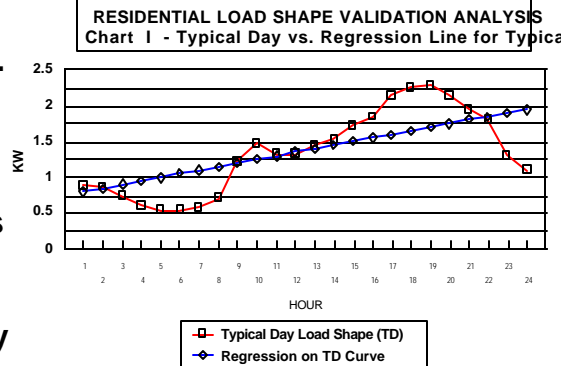


Applied Energy Group, Inc  
2000 AEIC Load Research Conference

19

## Validation Test - Pattern Recognition

- **Option 3: Use regression vs. time as the baseline for residuals**
- **Better reflects changes in use levels during the day**

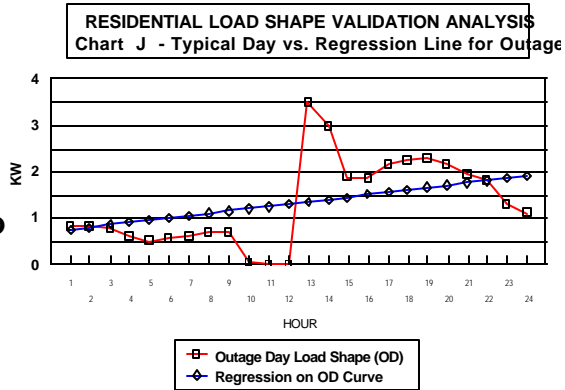


Applied Energy Group, Inc  
2000 AEIC Load Research Conference

20

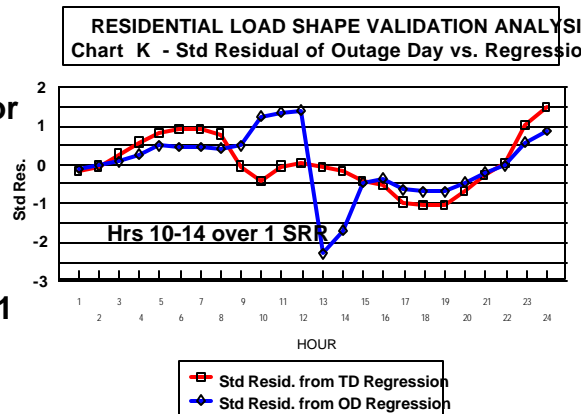
## Validation Test - Pattern Recognition

- Option 3: Regression line for Outage Day should be very similar to typical day



## Validation Test - Pattern Recognition

- Option 3: Standard Residuals for Typical Day vs. Outage Day
- Standard residuals > 1 indicate problems



## Validation Test - Pattern Recognition

- **Advantages of Pattern Recognition:**
  - Unbiased Test
  - Standardizing residuals accounts for different load levels - can develop base residual pattern from previous month's average load shape
  - Can be made computational and automated
- **Constraints of Pattern Recognition**
  - Useful only for short outages up to ~8 hours
  - Need additional test to validate use levels



Applied Energy Group, Inc.  
2000 AEIC Load Research Conference

23

## Editing Techniques

- **Smoothing**
  - Short-term outages (1-4 hours), use average value on either side for missing/invalid data
- **Shaping**
  - Short-medium periods, use typical shapes from previous period, with scaling option
- **Borrowing**
  - Medium to long periods, including whole days: “borrow” data from a previous period, similar in day of week, day type, weather



Applied Energy Group, Inc.  
2000 AEIC Load Research Conference

24

## Editing Techniques

- **Patterning**
  - Similar to Shaping, use standardized residual from prior period, scaled automatically from problem day use level
  - Appropriate for up to 8 to 10 hours within a day
  - Most valuable for outage days, where both outage and payback must be identified



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

25

## Editing Criteria

- “Rules” need to be consistent
- **Typical Rules:**
  - Smooth up to 4 missing/invalid intervals, including payback after outage
  - Shape or Pattern patch only up to 1/2 day
  - Borrow data to patch missing days, using same day type and similar weather day
  - Pattern data to patch missing days with external estimate of load level
  - Need an audit trail to document/undo

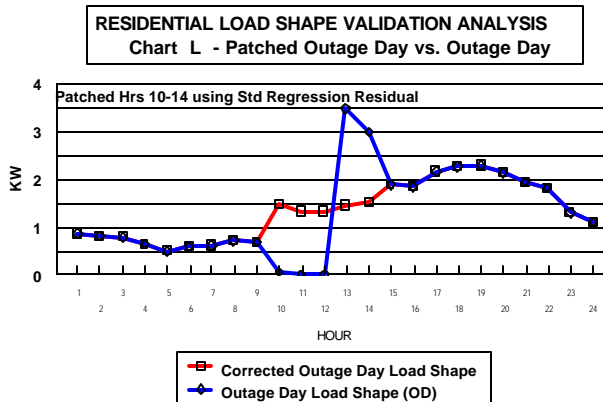


Applied Energy Group, Inc  
2000 AEIC Load Research Conference

26

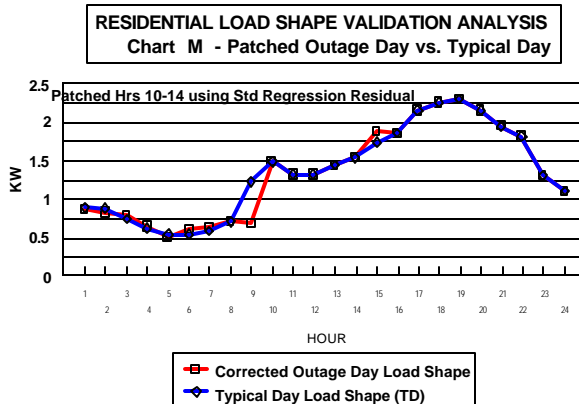
# Typical Edit Transaction

- Pattern Patch of Outage and Payback Period



# Typical Edit Transaction

- Pattern Patch of Outage vs. Comparable Typical Day
- Hours on either side of problems are still slightly off



# Validation & Editing Today

- Deregulated industry will increase the need for accurate, **QUALITY-CONTROLLED** Data
- Load Profiling and Market Settlement Process will require more automation as data must be processed daily and/or “on-the-fly” and expanded to supplier or system levels
- Many users for load data, including regulated utility, suppliers, energy service providers
- Applications: Load Profiling, rates, technology assessment and other traditional ones



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

29

# SUMMARY

- **SO MUCH DATA, SO LITTLE TIME!**
- Deregulation will require automated data quality control: Validation and Editing
- Risk of expanding data with unseen problems is great - and avoidable!



**BEFORE**

Load Research Man says: “Don’t be a caveman like me!”



**AFTER**



Applied Energy Group, Inc  
2000 AEIC Load Research Conference

30