Learning to Predict Channel Stability Using Biogeomorphic Features

Stephanie L. Moret

Louisiana State University

William T. Langford

National Center for Ecological Analysis and Synthesis, UC Santa Barbara

Dragos D. Margineantu

The Boeing Company

Excess Channel Instability can adversely affect Streams

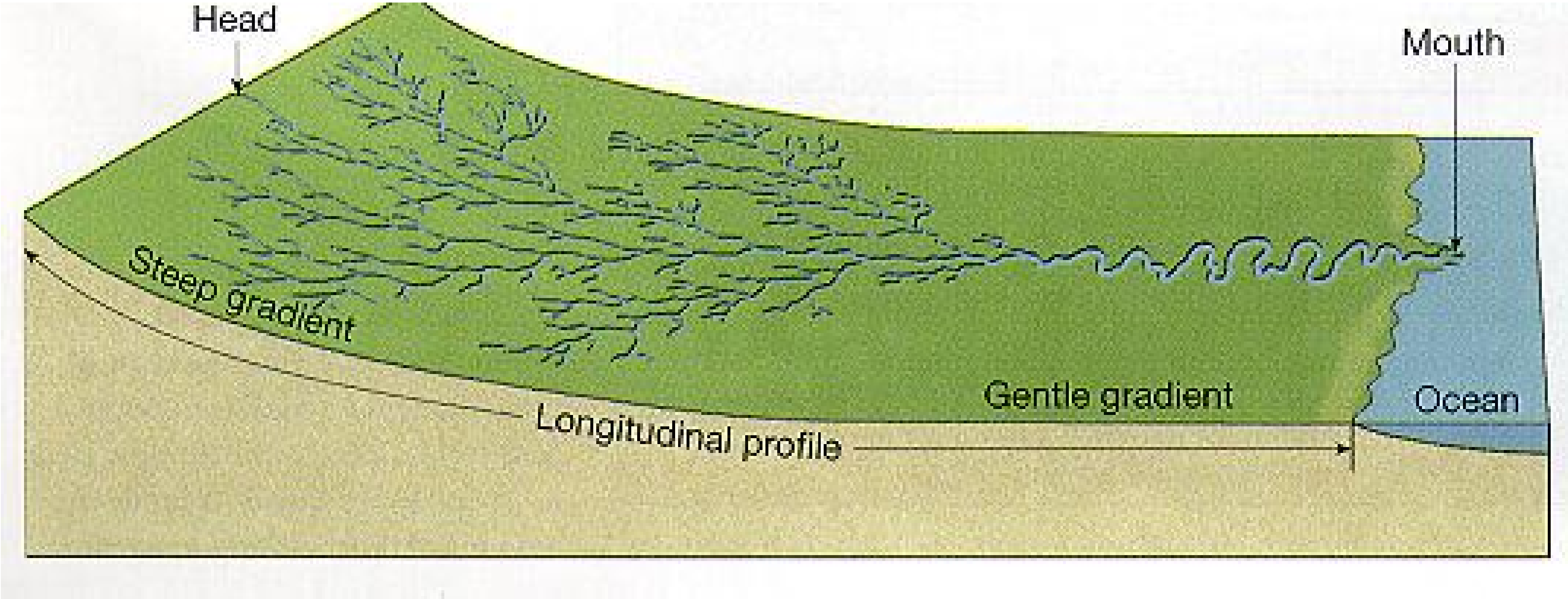
- **Aquatic and Riparian Habitat**
  - **Sediment as non-point source pollutant**
    - **Smothers salmon redds (egg nests)**
    - **Increases temperature, causing respiratory distress**

- **Degrades bed and undercuts banks**
  - **Killing riparian vegetation**
    - **Lose bird habitat**
    - **Removes shade => warmer water**
  - **Altering channel shape and function**
    - **Flat, wide => warm, mossy => bass**
      - **Bass prey on salmon fry**
    - **Deep, narrow => faster, colder => salmon**

- **Property and Life Loss**
  - **Undermine pipes and bridge supports**
  - **Bank failure**
  - **Instigates landslides by cutting toe of slope**

Channel Stability

Ability of stream, over time,
to transport flows and sediment
without changing geomorphic character of river



Some Channel Stability Indicators

- Upper Banks
  - **Landform slope**
  - **Vegetation**
- Lower Banks
  - **Bank Rock Content**
  - **Cutting / Deposition**
- Channel Bottom
  - **Particle Packing**
  - **Scouring**
  - **Deposition**

Stable

Well vegetated, low slope

Unstable

Poorly vegetated, cut bank, deposition bars

Rapid Assessment Protocols

- Detailed field studies
  - Expensive
  - Time consuming
- RAPs
  - Cost effective
  - Multiple sites in a field season
  - Extensively used by government agencies

Channel Stability RAP

- Stream Reach Inventory and Channel Stability Evaluation (Pfankuch, 1978)
- Developed in 1978 by the USDA FS Rocky Mountain Station
- Used in 60% of US National Forests

Channel Stability Indicator Item	excellent	good	fair	poor
Upper Banks:				
Landform Slope	2	4	6	8
Mass Wasting	3	6	9	12
Debris Jam Potential	2	4	6	8
Vegetative Bank Protection	3	6	9	12
Lower Banks:				
Channel Capacity	1	2	3	4
Bank Rock Content	2	4	6	8
Obstructions and Flow Deflectors	2	4	6	8
Cutting	4	8	12	16
Deposition	4	8	12	16
Channel Bottom:				
Rock Angularity	1	2	3	4
Brightness	1	2	3	4
Consolidation/ Particle Packing	2	4	6	8
Bottom Size Distribution	4	8	12	16
Scouring Deposition	6	12	18	24
Aquatic Vegetation	1	2	3	4

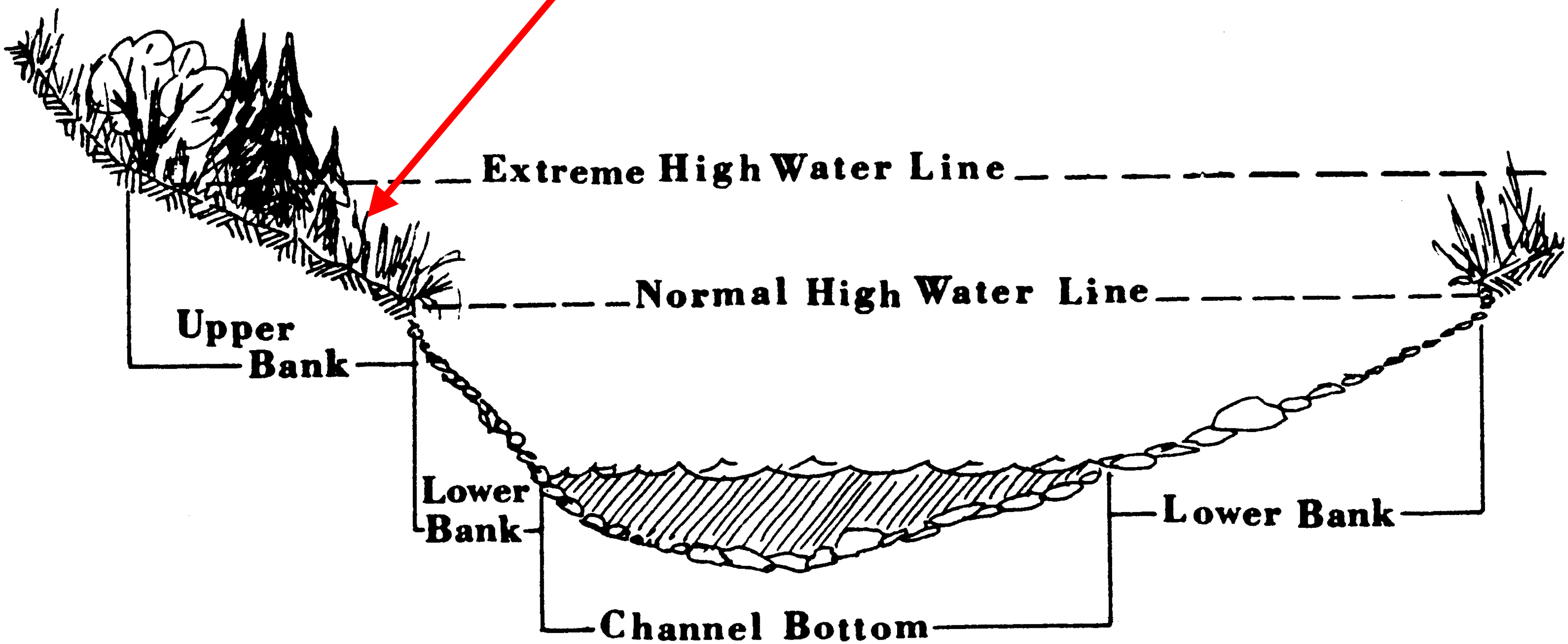
Sum =

TOTAL

REACH

SCORE

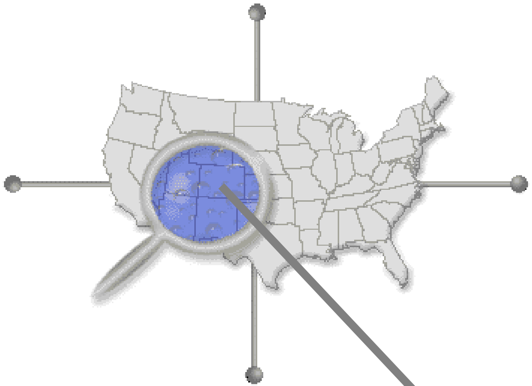
Channel Stability Indicator Item	excellent	good	fair	poor
Upper Banks:				
Landform Slope	2	4	6	8
Mass Wasting	3	6	9	12
Debris Jam Potential	2	4	6	8
Vegetative Bank Protection	3	6	9	12
Lower Banks:				
Channel Capacity	1	2	3	4

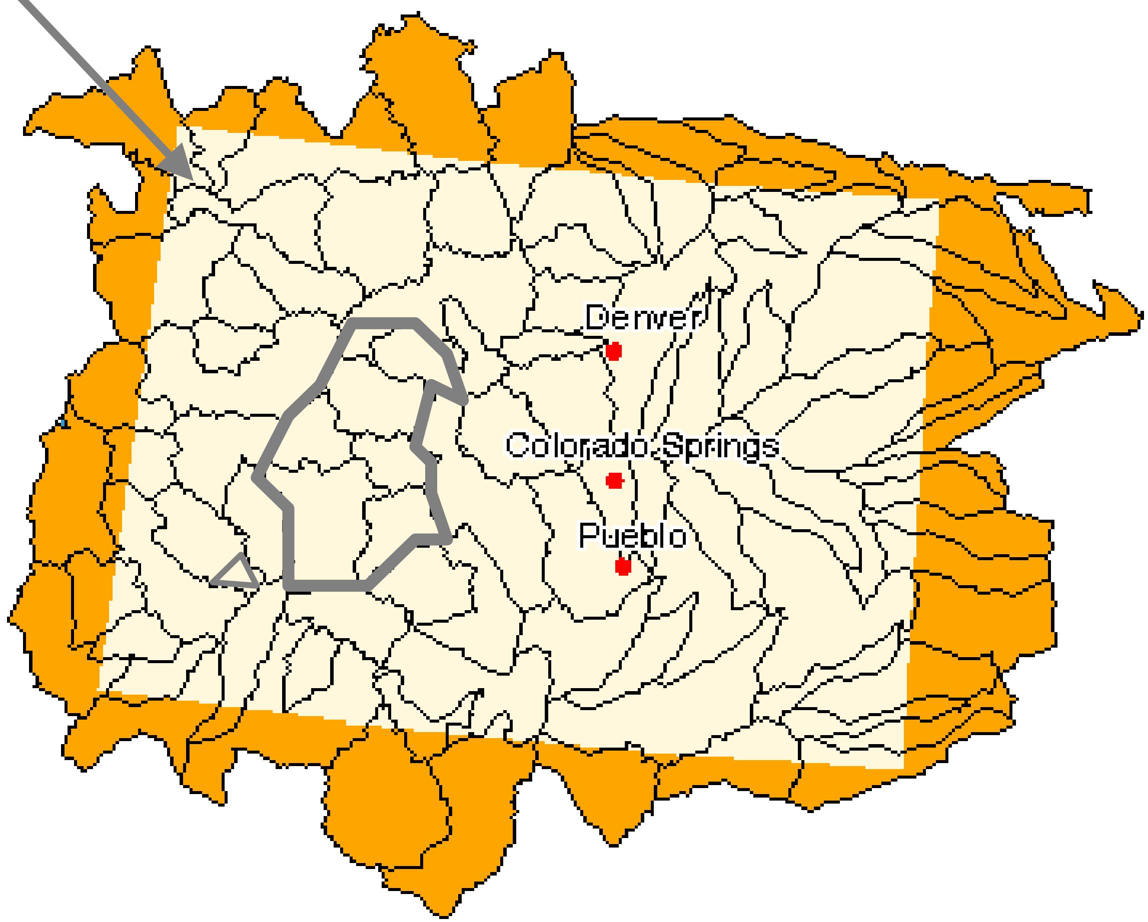


RAP problem

- Many streams still go unmonitored
- Solution: Learn to predict RAP outcome
  - Collect Channel Stability training data using RAP
  - Collect GIS attributes for each site
  - Create cost model

Study Site:
Upper Colorado River Basin





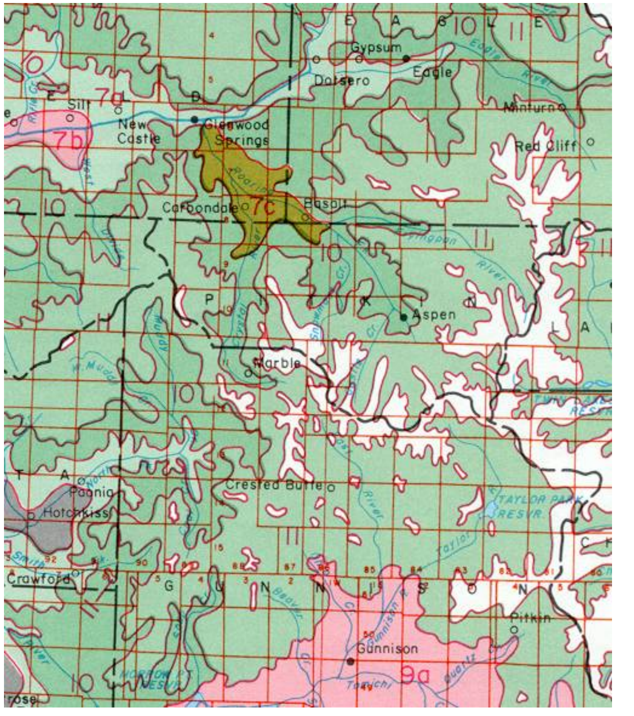
60 high-elevation 3^{rd} & 4^{th} order streams

Input Features

- Stream Gradient
- Channel Sinuosity
- Elevation
- Precipitation

- Geology
  - Igneous
  - Metamorphic
  - Sedimentary
  - Quaternary Alluvium
- Land Use / Land Cover
  - Forested Land
  - Rangeland
  - Pasture and Cropland

GIS Source Data



- USGS Geologic Map of Colorado
- USGS 7.5 minute topo maps
- NSDA NRCS Colorado Annual Precipitation Map
- EPA BASINS, Region 8
  - GIRAS- Anderson Level II Land Classification

Predicting scores by regression too difficult

- Convert to classification problem
  - Threshold scores
- Use cost-based learning
  - Minimize cost instead of percentage error
    - Since some errors much more important

System Costs

	Actual Stable (-)	Actual Unstable (+)
Pred. Stable (-)	TN: no cost	FN: Huge cost (lives, bridges, roads, prop., hab.)
Pred. Unstable(+)	FP: 1 day labor	TP: 1 day labor + Full Cost of Remediation

First cut results

- None of these algorithms can recognize all of the unstables
  - Threshold linear model
  - Logistic regression
  - Decision trees
- One problem is poor probability estimates

Solution: Improve probability estimates

- Techniques for improving esimates
  - Probability Estimation Trees
  - Lazy Option Trees
  - Bagging

Bagged Probability Estimation Trees

(Provost & Domingos, 2000, 2002)

Probability Estimation Trees (PETs) are:

- decision trees (e.g. C4.5)
- no pruning/collapsing
- Laplace correction at leaves

$$P(y = k \mid \text{leaf}) = \frac{n_k + 1/K}{N + 1}$$

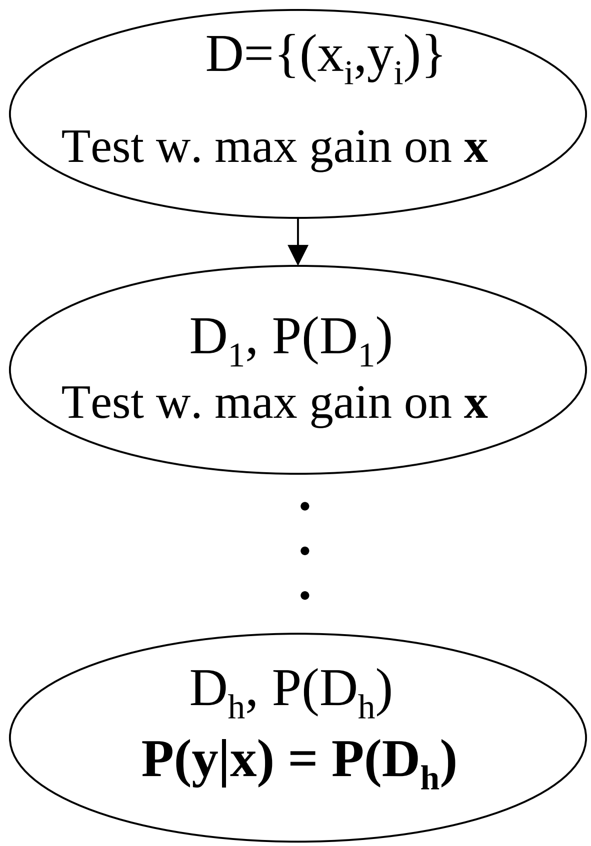
Bagged PETs (**B-PET**s).

- average estimates across a bag of trees.

Lazy Decision Trees

Lazy Decision Trees:

For **each** previously unseen **instance** w = (**x** , **y**) grow a tree:



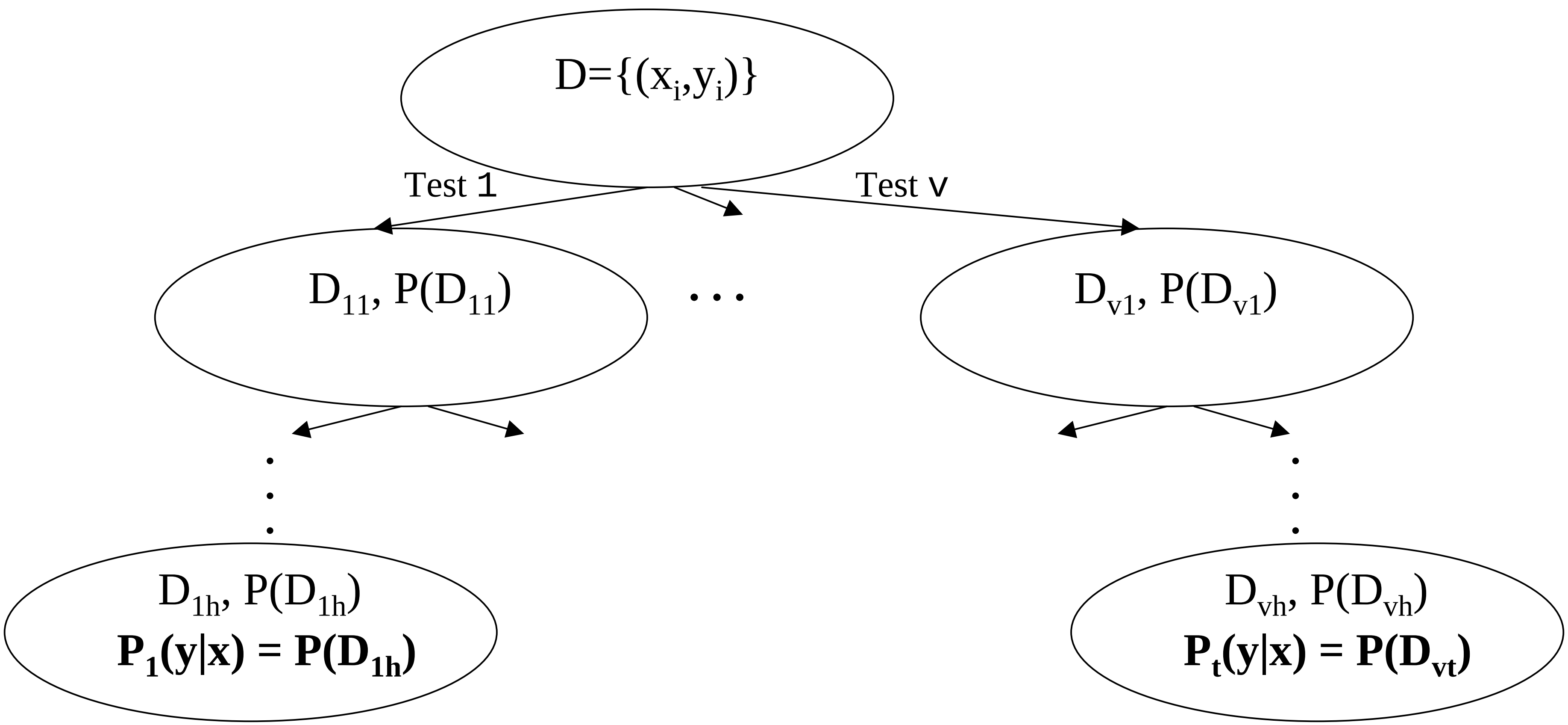
+ Options at Nodes

=> Lazy Option Trees

(Margineantu & Dietterich, 2001, 2002)

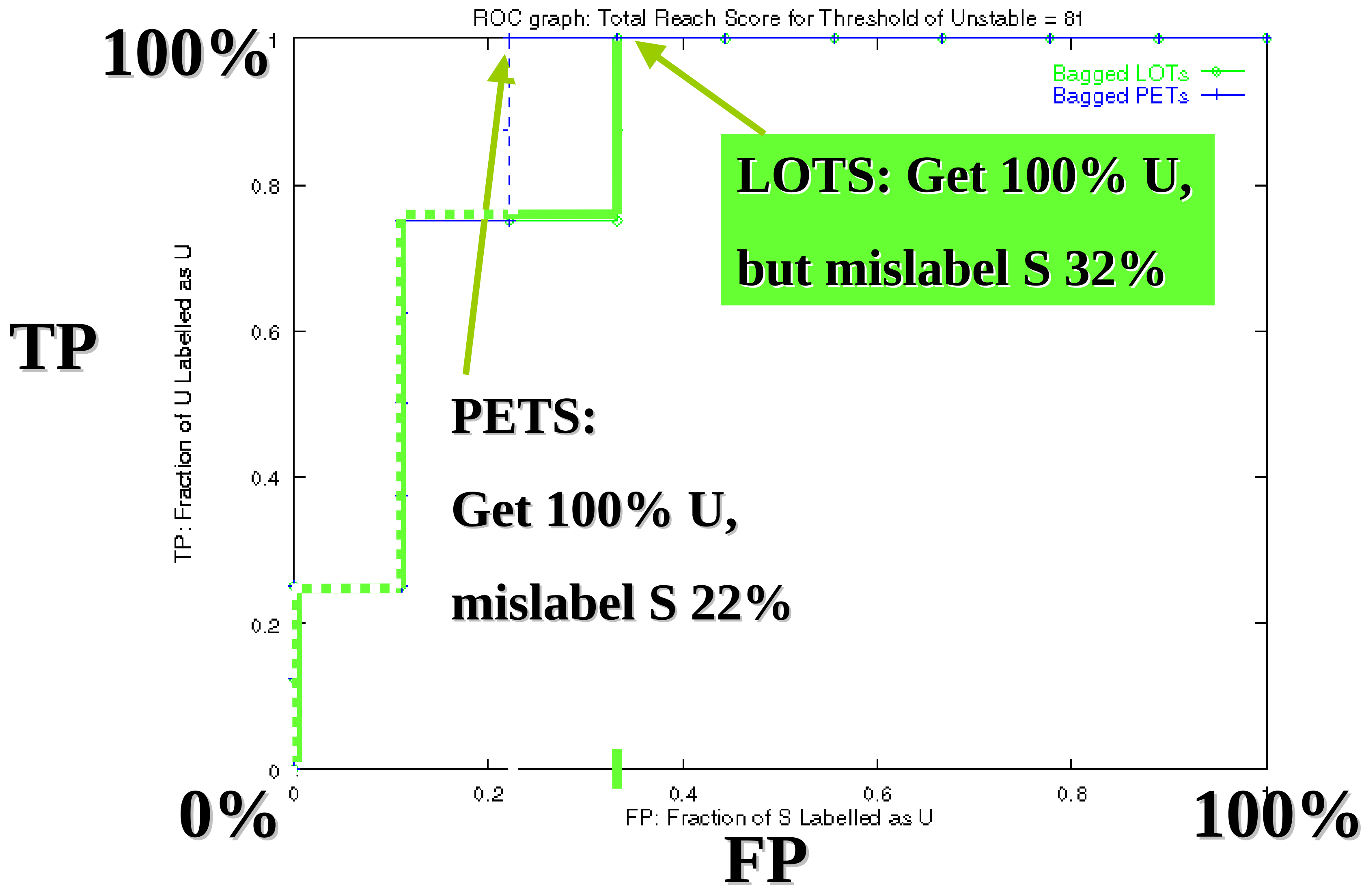
Lazy Option Decision Trees (**LOT**):

For **each** previously unseen **instance** w = (**x** , **y**) grow a tree:



Results for one threshold

Threshold: 81



Results:
Lowest error fractions where all unstables are recognized

θ	B-LOTs	B-PETs	Fraction Unstable Sites in Dataset
81	0.33	0.22	0.49
84	0.60	0.40	0.45
86	0.72	0.45	0.42
88	0.73	0.55	0.36
90	0.47	0.58	0.29
93	0.50	0.75	0.25
95	0.28	0.50	0.20

Discussion

- Only B-PETs and B-LOTs could find all unstables
- PETs and LOTs each better at different thresholds
  - B-LOTs better when unstables are rare
- Hard problem (often only about 50% right)
  - But still at least 50% cost savings for agency
  - And “perfect” safety

Conclusion

- Cost-sensitive approach
  - with good probability estimation enables solution
- May be useful for any RAP situation
  - Especially since so much GIS data available
- Future work: Simple to attach to a GIS