What data quality do we need and why?

Pravin Varaiya, University of California, Berkeley

TRB ITS Committee mid-year meeting on real-time traveler information systems, April 15, in Irvine, California

1

Outline

1. Two propositions
2. The magic number is 25
3. Can cell phones, LPR, ETC help?
4. What is travel time reliability?
5. Sustainable traffic management

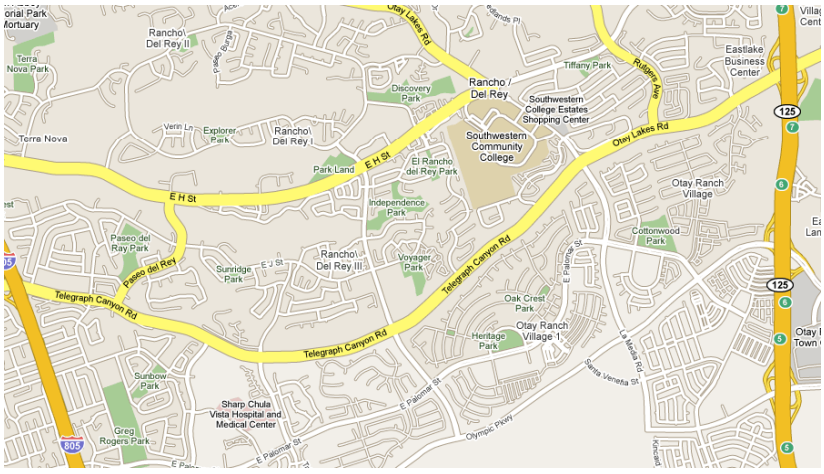
2

Two propositions

- If you didn't measure how your streets performed 5 minutes ago, don't expect to manage them well now.
- An inaccurate travel time estimate is worse than no estimate.

3

A problem

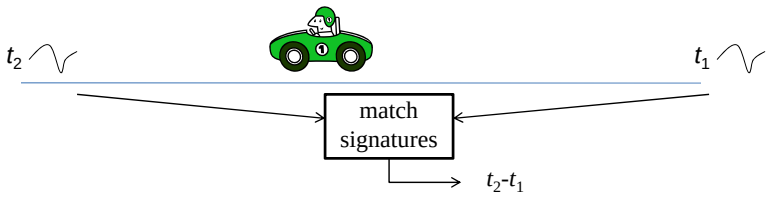


- Estimate the travel time on Telegraph Canyon Road.

4

Need to re-identify vehicles

- Travel time across a road segment can only be obtained from individual vehicle travel times. Point speeds (loops, microwave) are useless because of stop-and-go movement.



Individual travel times need signature matching based on:

License plate, toll tag – expensive, privacy

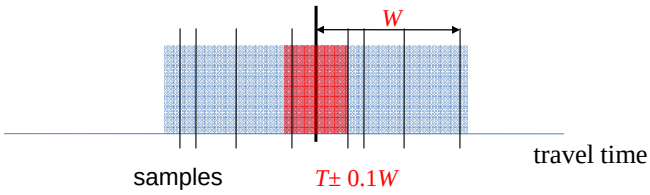
GPS, Bluetooth phone – 'cheap', privacy, coverage, errors

Magnetic signature – anonymous, full coverage

5

The magic number is 25

- How many individual vehicle travel time samples does one need?



- To estimate the median travel time $T \pm 10\%$ spread with 68% confidence you need $N = 25$ samples; for 95% confidence you need $N = 100$.

6

Sample size consequences

- A street with flow of V vph needs sample penetration rate of $p = 25/V \times 100\%$
  - If $V = 400$, $p = 6\%$
  - If $p = 2\%$, you must wait 3 hours for a reliable estimate
- Suppose 10,000 GPS (or Bluetooth)-equipped vehicles in Bay Area record their trip times. What kind of penetration is achieved?

  With 4.7m Bay Area vehicles making 18m trips daily, 10,000 vehicles making 38k trips achieve $p = 0.2\%$. The resulting estimates will be unreliable.

7

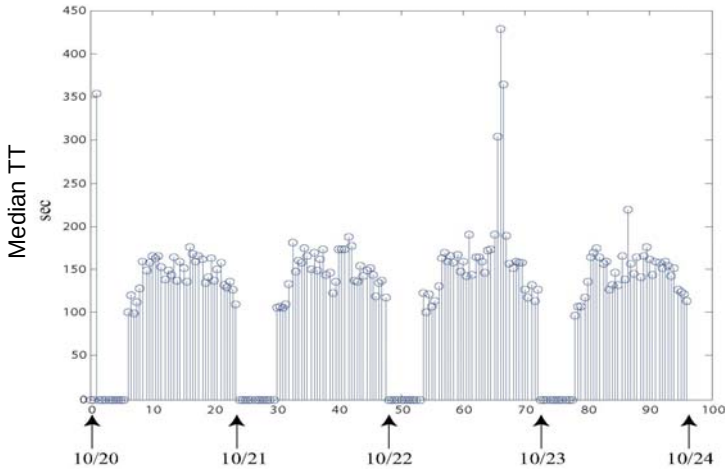
Further consequences

- The sample size requirement of ($V = 400$, $p = 6\%$) , ($V = 100$, $p = 25\%$) assumes that travel times do not change over the course of one hour.
- If during congestion, conditions change in (say) 30 min, the sample size requirements will be doubled ($V = 200$, $p = 12\%$), ($V = 50$, $p = 50\%$).
- Of course, travel time estimates are most useful when conditions change.

8

An example

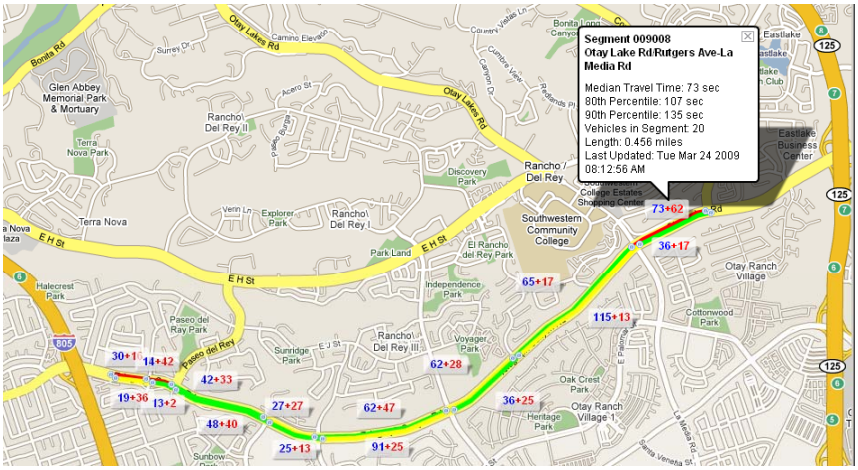
- I-880 acciden t on 10/22/2008 shows dramatic change in travel time within 30 min on San Pablo (8 miles away)



9

What is travel time reliability?

- Giving a single number for TT is meaningless without knowing its variability, e.g. median and 80th percentile.



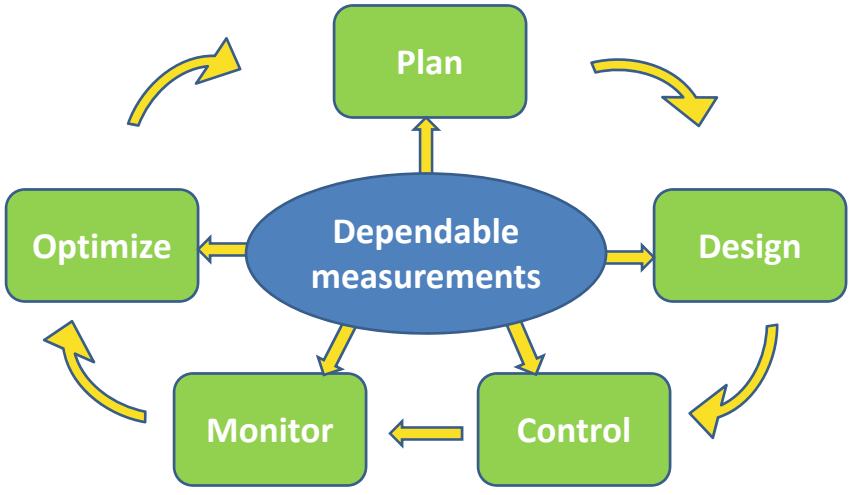
10

A caution about arterial management

- Knowledge of travel time distribution by itself is insufficient for arterial management.
- You also need to know arterial performance: volume, delay, efficiency of intersections, and information to improve signal control.
- This requires dependable measurements:
  - Dependability = coverage × accuracy

11

Sustainable urban traffic management



12

Conclusion

- Matching LP, ETC or cell-phone numbers cannot achieve desired sample size today.
- Matching magnetic signatures achieves required penetration rate.
- Travel time reliability is essential for estimate to be trusted.
- For urban traffic management one needs a dependable measurement system.

13