Visualization and Analytic Methods for the Tracking of Birth Outcomes and Traffic Exposures

• Paul English, PhD MPH, Principal Investigator
• Eric M. Roberts, MD PhD, Pilot Project Manager
• Craig Wolff, MS Eng, IT/GIS Manager
•Rusty Scalf, GIS Specialist

Environmental Health Investigations Branch
California Department of Health Services

• Funding provided by Centers for Disease Control and Prevention, the National Cancer Institute, and NIEHS

Overview

- Needs of Visualization for Tracking
- Mapping Disease Rates
  - Failure of mapping in discrete areas
  - Density estimation
  - Strengths and Weaknesses
- Modeling of Traffic-Exhaust Pollution
  - Methods: Cost-Benefit Analysis

Needs of Data Visualization for Tracking

- We need a system that is continuous and ongoing
- We need web-based tools that allow public access and ability to interface with data
- We need to preserve data confidentiality and privacy
- The system is not complete until those who need information
  - Know the information exists
  - Know where to find it
  - Know what it's good for
  - Are able to access and interpret it

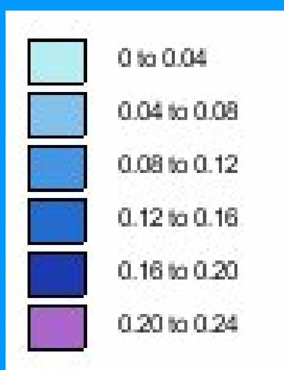
Mapping Disease Rates

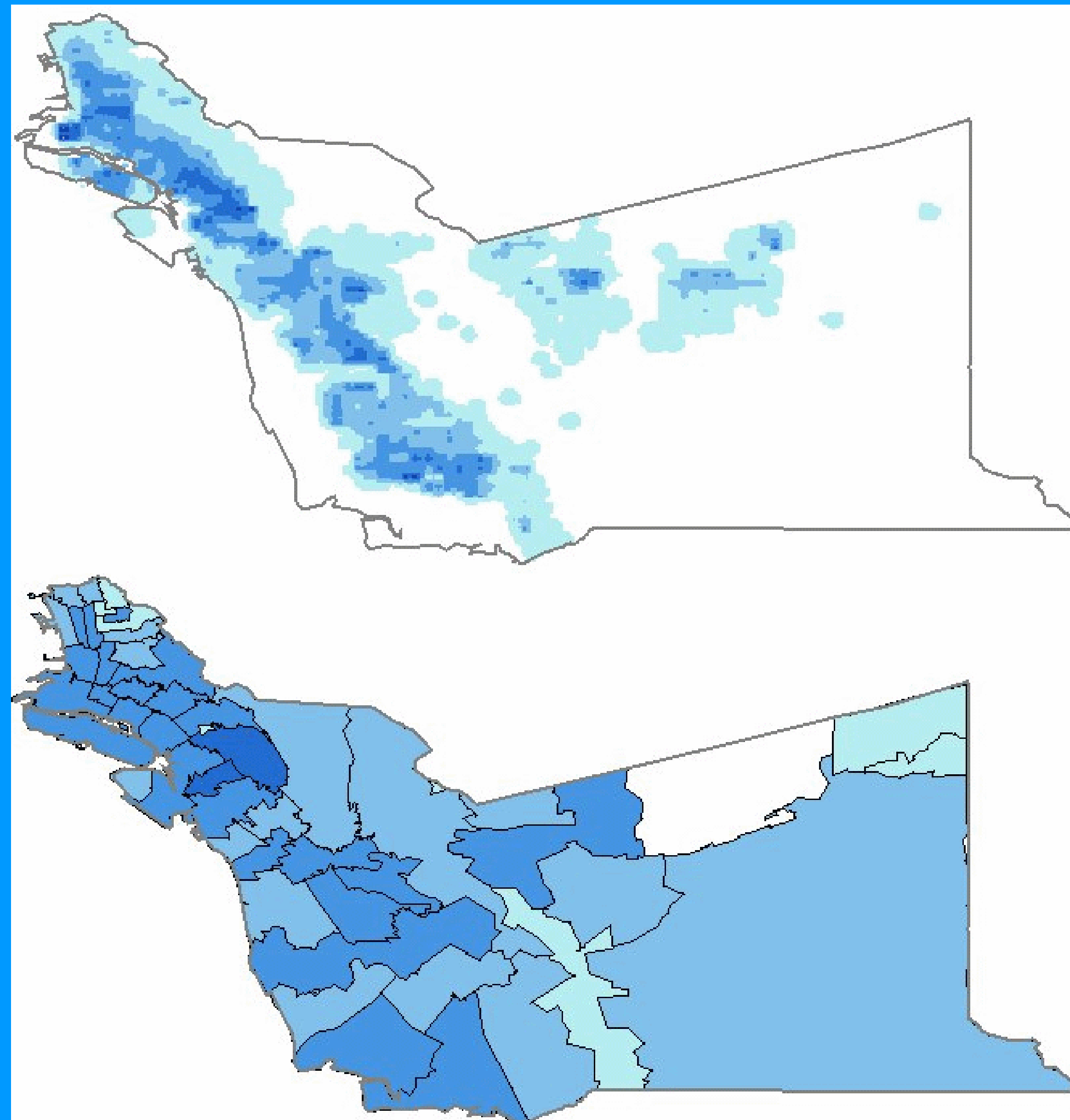
- Failure of mapping rates in discrete areas
  - Sample size problem: law of small numbers
    - Instability of rates with small denominators
    - As areas get smaller, variability increases
  - Visualization fails
  - Political boundaries change over time (e.g. ZIP codes)

On-going monitoring and dissemination of information on the distribution of environmentally related disease

Pilot Project 1

Pre-term birth rate, Alameda County, 2001 (By density estimation and by zip)





Density Estimation: Strengths

- Methods of Openshaw,
  Rushton

Produce continuous surface of rates

Preserves data confidentiality

More accurately reflect reality

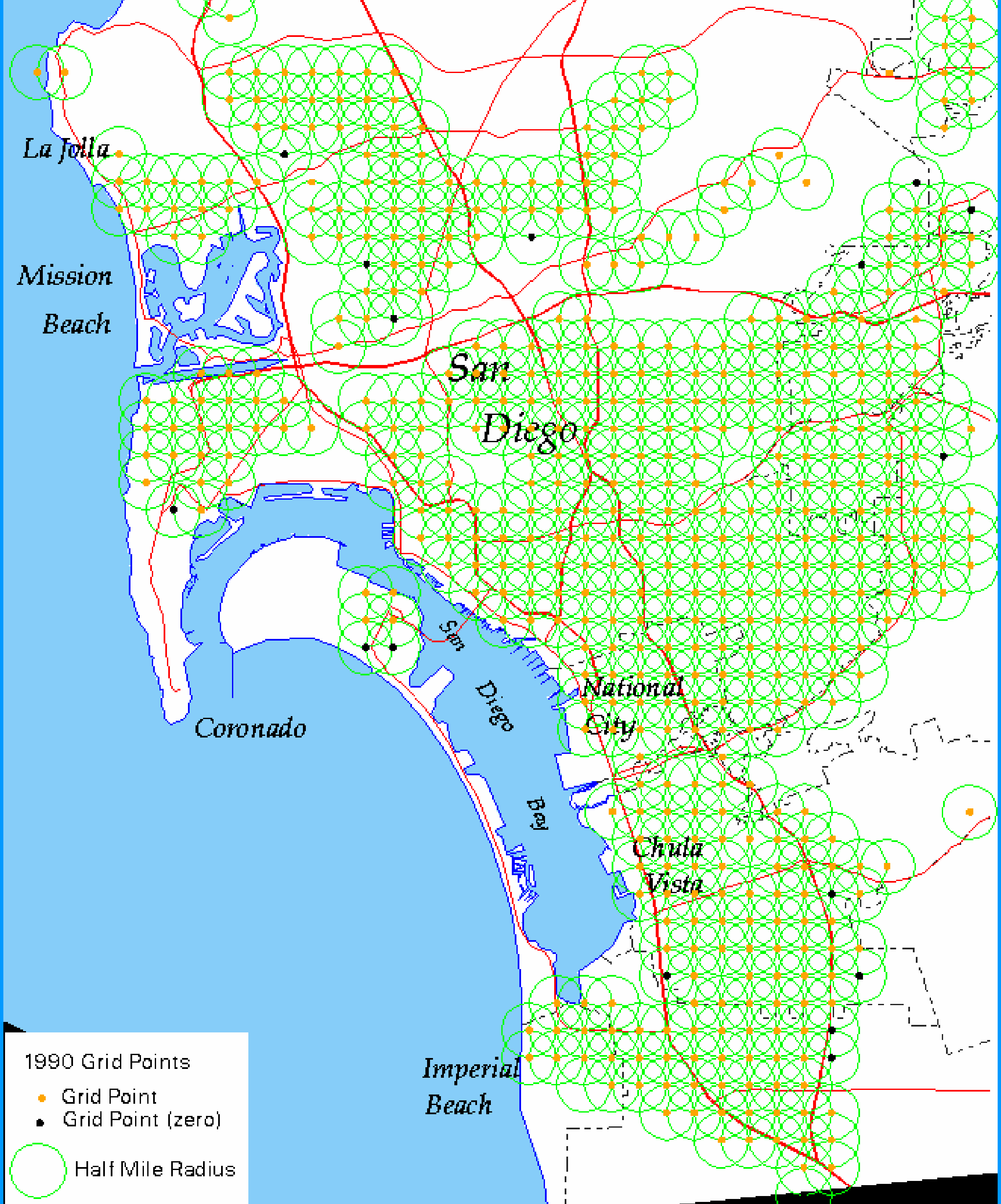
Data Restrictions

- Restricted births to:
  - completed birthweight
  - geocoded address
  - downtown San Diego and nearby areas
  - compatible birthweight and gestational age

  - 16,385 births 1980
  - 24,274 births 1990

Spatial Distribution Method

(1) Generated Uniform Grid 0.5 miles apart (spatial filters)

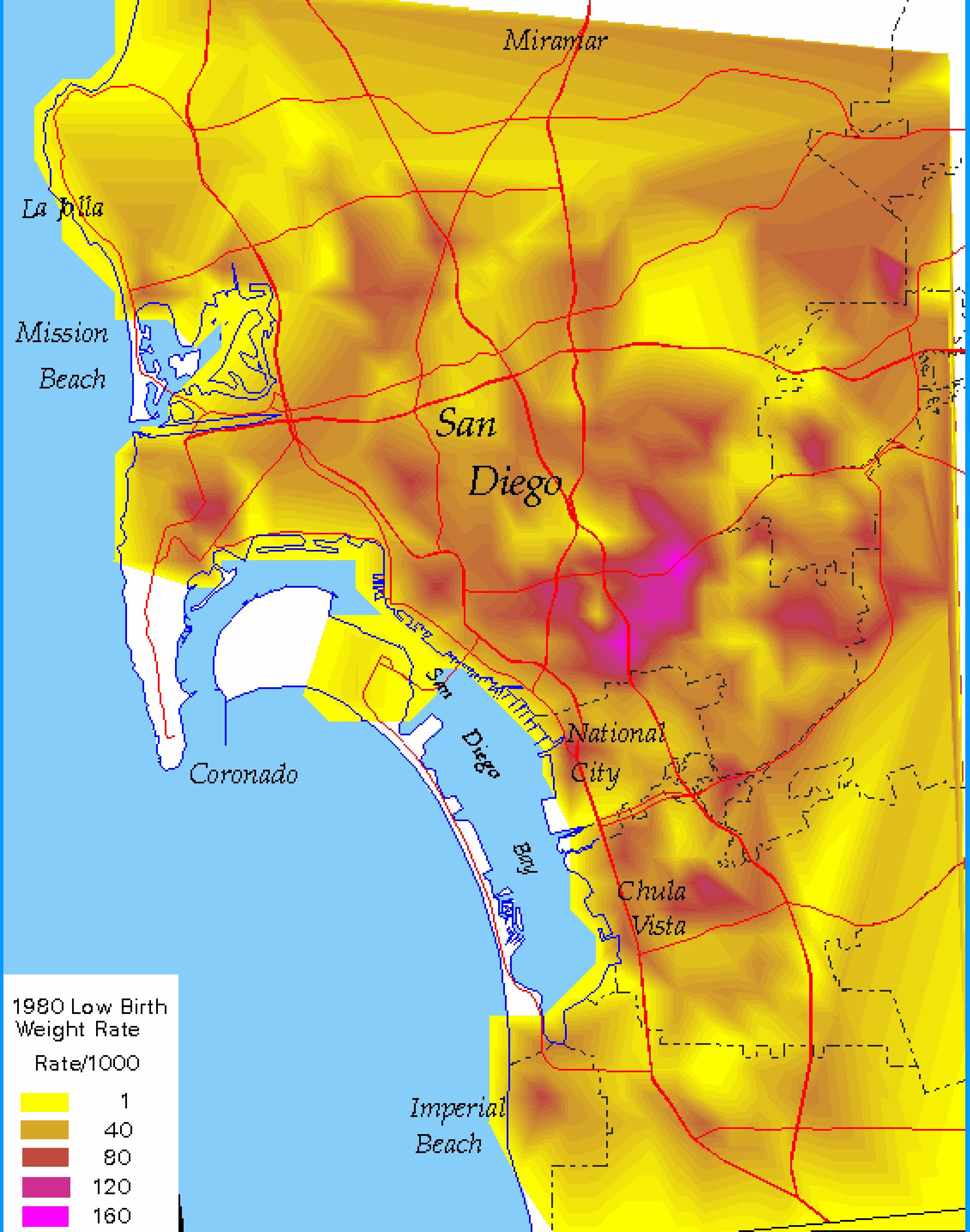


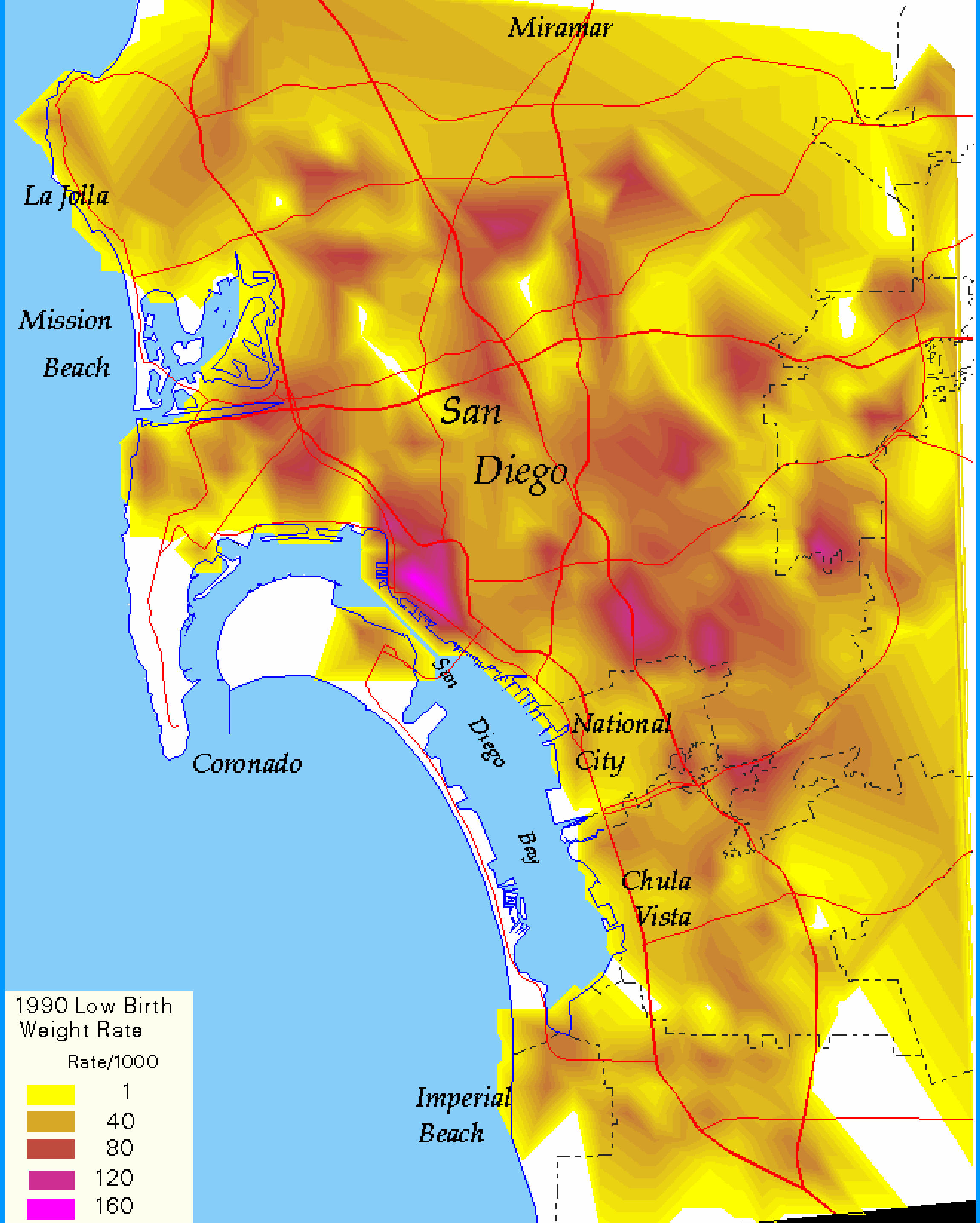
La Jolla
Mission Beach
San Diego
San Diego Bay
National City
Coronado
Chula Vista
Imperial Beach
1990 Grid Points
Grid Point
Grid Point (zero)
Half Mile Radius

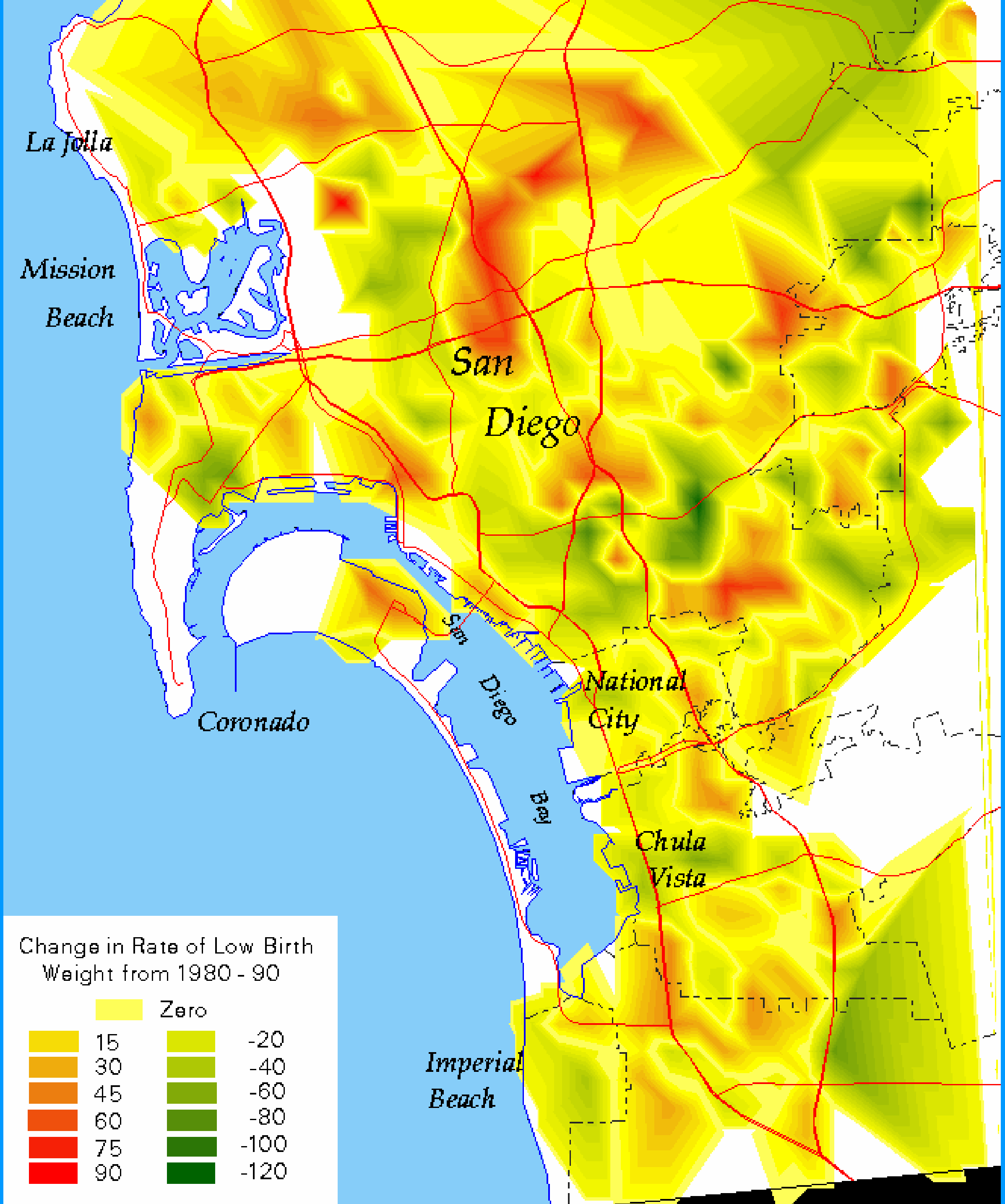
Spatial Distribution Method

(2) Identified all births within a 0.5 mile spatial filter (min. of 40 births to compute rate)

3) Compute LBW rates and made contour maps


Miramar
La Jolla
Mission Beach
San Diego
Coronado
San Diego Bay
National City
Chula Vista
Imperial Beach
1980 Low Birth Weight Rate
Rate/1000
1
40
80
120
160

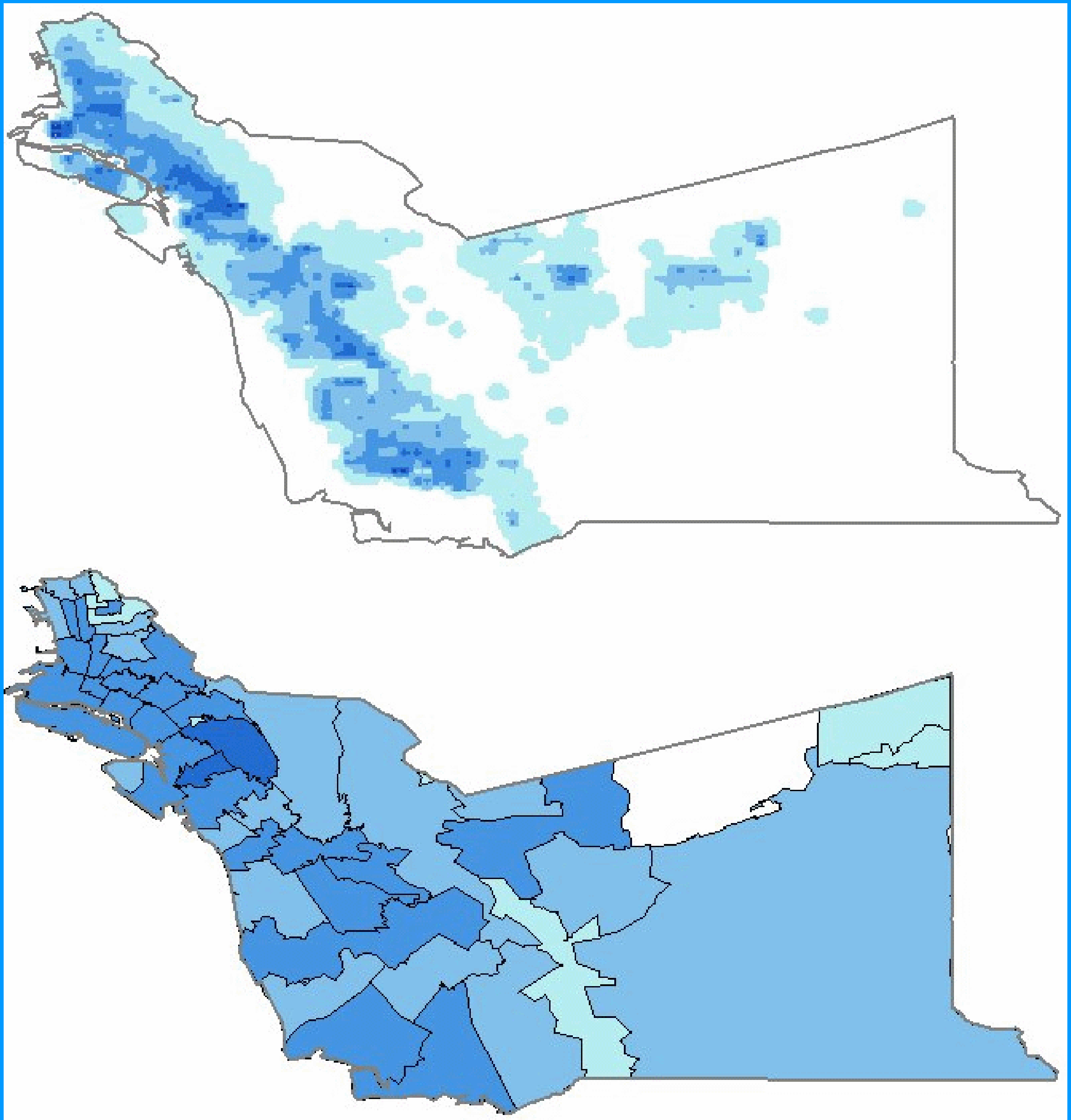

Miramar
La Jolla
Mission
Beach
San
Diego
National
City
Coronado
San
Diego
Bay
Chula
Vista
Imperial
Beach
1990 Low Birth
Weight Rate
Rate/1000
1
40
80
120
160


La Jolla
Mission
Beach
San
Diego
San
Diego
Bay
National
City
Coronado
Chula
Vista
Imperial
Beach
Change in Rate of Low Birth
Weight from 1980 - 90
Zero
15
30
45
60
75
90
-20
-40
-60
-80
-100
-120

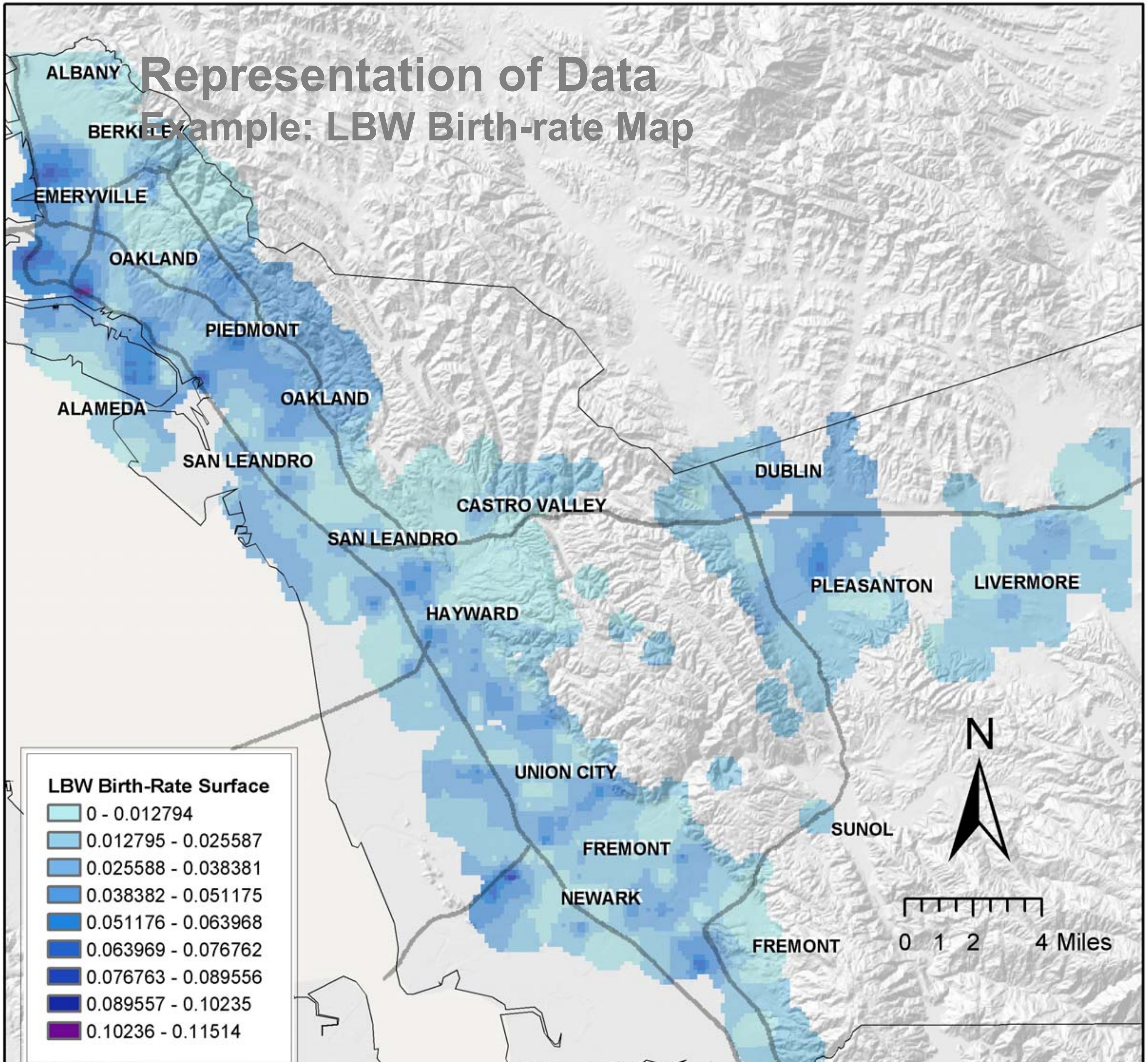
On-going monitoring and dissemination of information on the distribution of environmentally related disease

Pilot Project 1

Pre-term birth rate, Alameda County, 2001 (By density estimation and by zip)



Representation of Data
Example: LBW Birth-rate Map



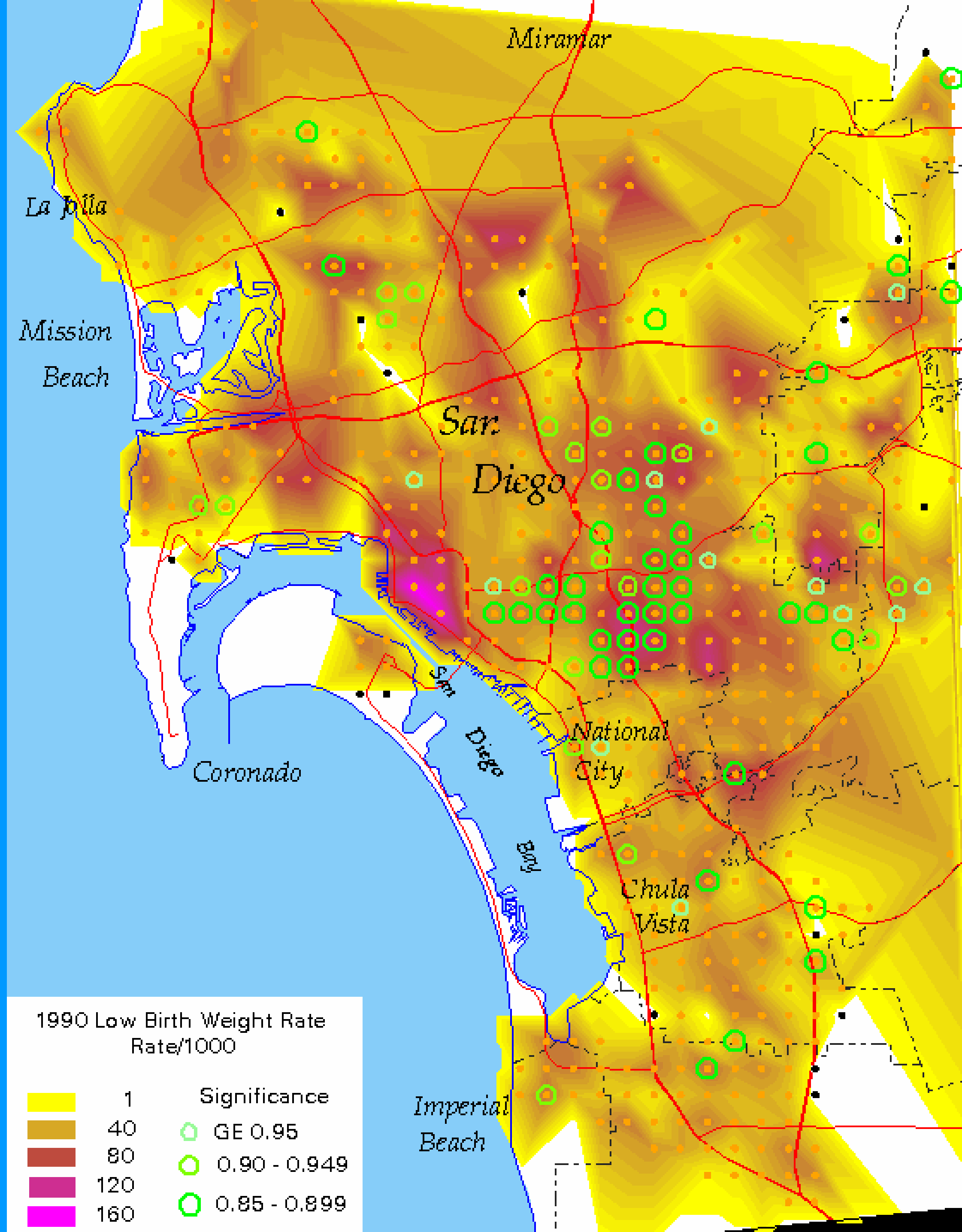
Statistical Significance (Monte Carlo Simulations)

- Assign each birth equal probability of being a LBW birth
- Assign a random number (1-1000) for each birth
- Classify each birth as LBW or non-LBW
- Compute LBW rate at each grid point 1000 times

Statistical Significance (Monte Carlo Simulations)

- Compute the % of simulated rates which are less than the observed rate.

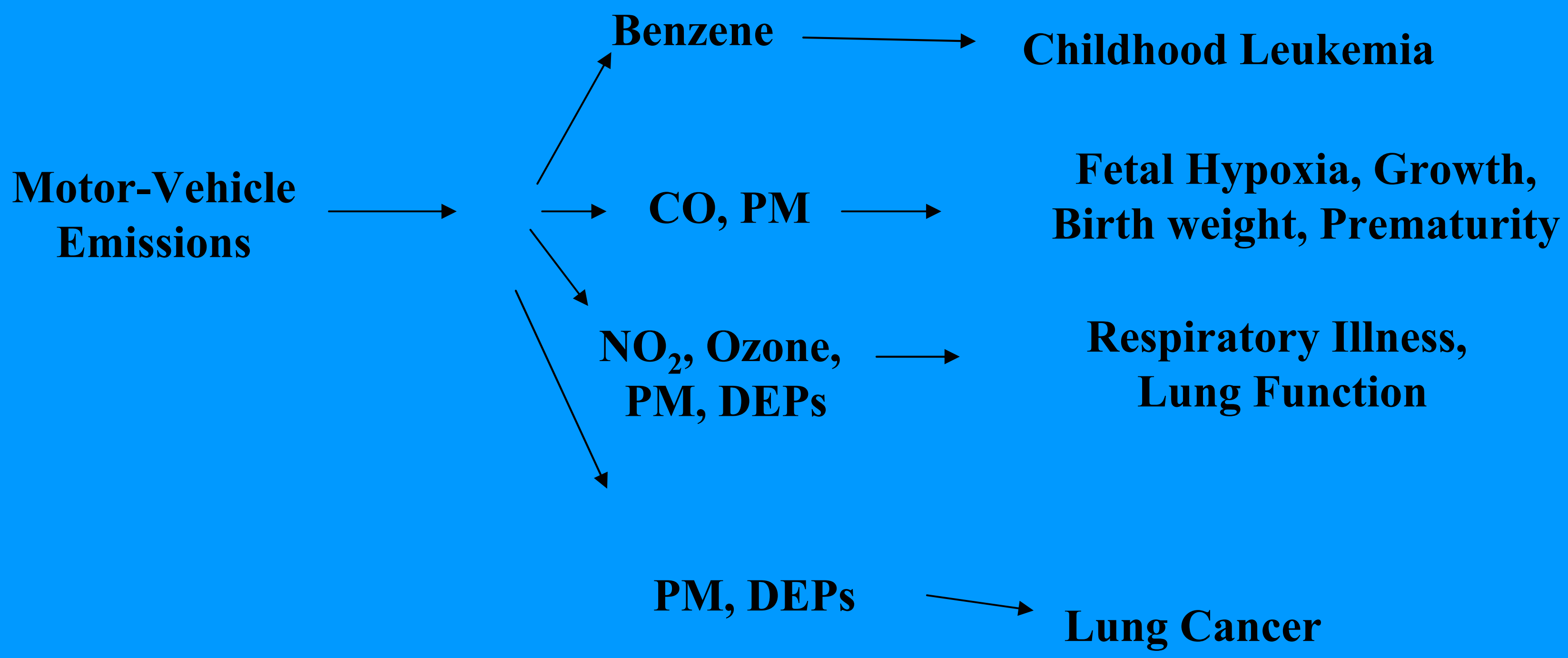
- Make map of statistical significance


Miramar
La Jolla
Mission Beach
San Diego
Coronado
San Diego Bay
National City
Chula Vista
Imperial Beach
1990 Low Birth Weight Rate
Rate/1000
1
40
80
120
160
Significance
GE 0.95
0.90 - 0.949
0.85 - 0.899

Density Estimation: Weaknesses

- Need for geocoded denominator
- Need for sensitivity analysis:
  - size of filter needs to change in relation of density of health events?
- Spatial autocorrelation an issue in analysis: Spatial regression

Health Effects Associated with Motor-Vehicle Emissions



Modeling of Traffic Exhaust Pollution

- Proximity Analyses
- Dispersion Models
- Land-Use Regression
- Integrated-Meterological Models

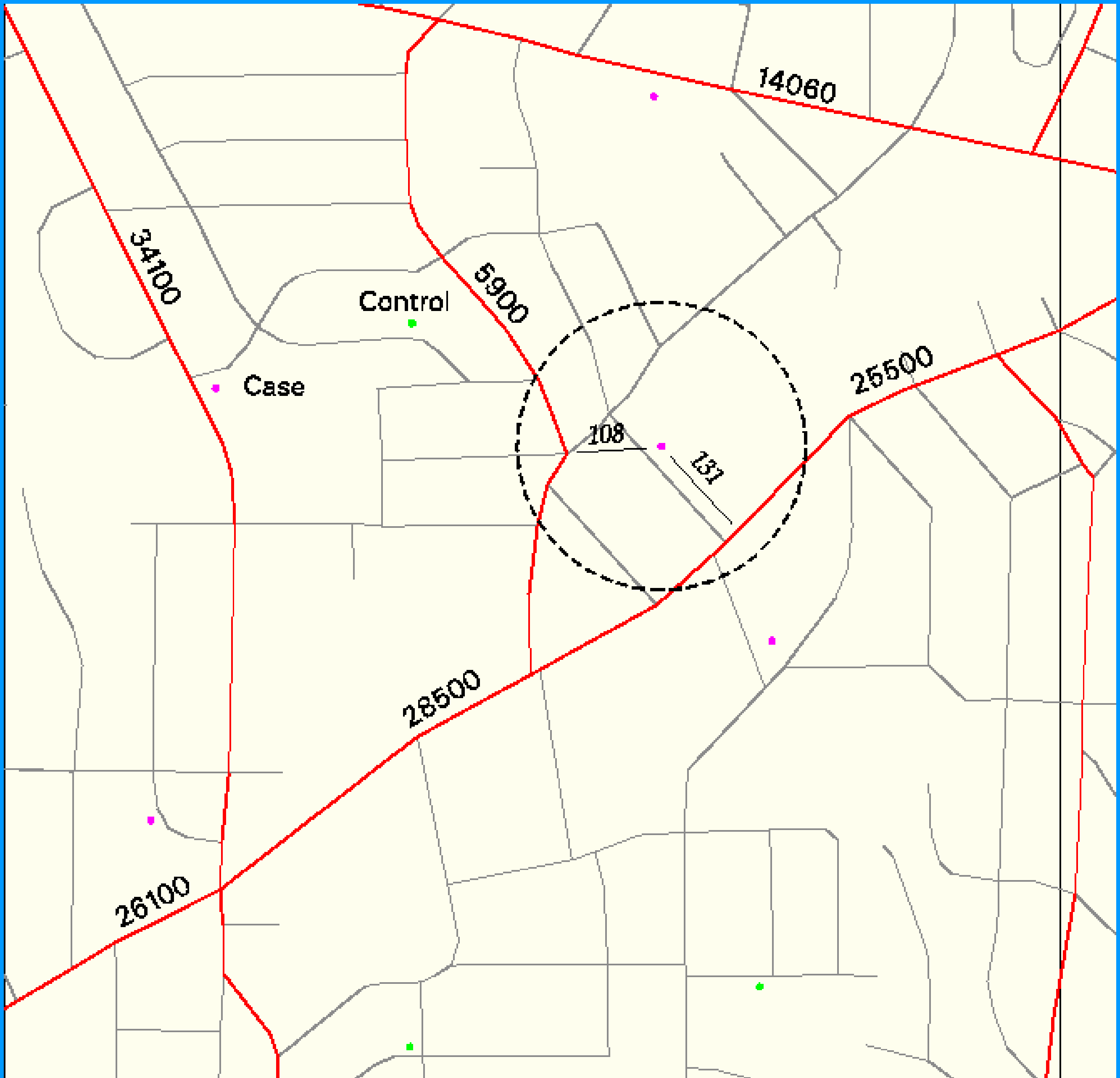
Simple to Complex Measures of Traffic-Related Pollutants

	Measure	Limitations:
Complex ↑	More complex modeling (e.g. ADMS- Urban)	Computationally complex, cost
	GIS w/ dispersion modeling GIS w/ regression-based models	More data inputs (e.g. meteorology, building configurations, emissions, etc.)
	Residence near fixed air monitors	Assumes homogenous exposure
	Distance-weighted traffic volume	Misclassification (e.g. building heights)
Simple	Neighborhood vehicle density	Misclassification (e.g. wind)

Proximity Analysis

- Strengths:
  - Easy to use (adapt with Gaussian weights)
    - Distance of residence to road correlates well with personal and ambient NO_2 monitoring (Rijnders,et al 2001)
  - Weaknesses:
    - Exposure misclassification likely without wind direction data
    - Does not model actual level of pollutants

Proximity Analysis

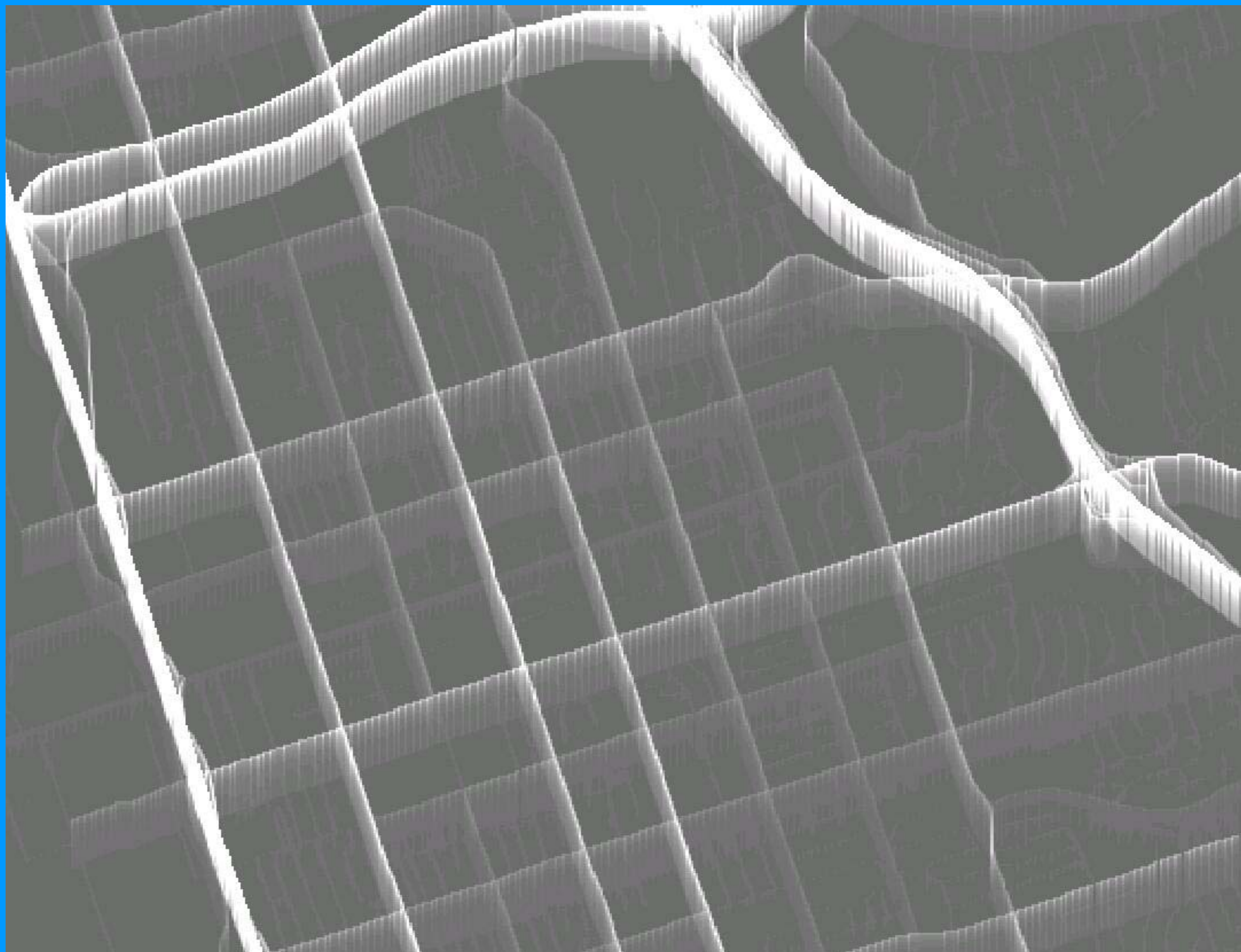


Dispersion Models (e.g. CALINE)

- Strengths:
  - More accurately measuring dispersal of pollutants
  - Model actual pollutant levels

  - Weaknesses:
    - Is gaussian plume realistic model?

Surface showing dispersal of hydrocarbons - light (high exposure) to dark (low exposure)
each pixel has grams hydrocarbons per 10 meters



Land-use regression (e.g. Briggs)

- Strengths:
  - Use land use, met data, DEMs, traffic to predict pollutant concentrations
  - Easily obtained data

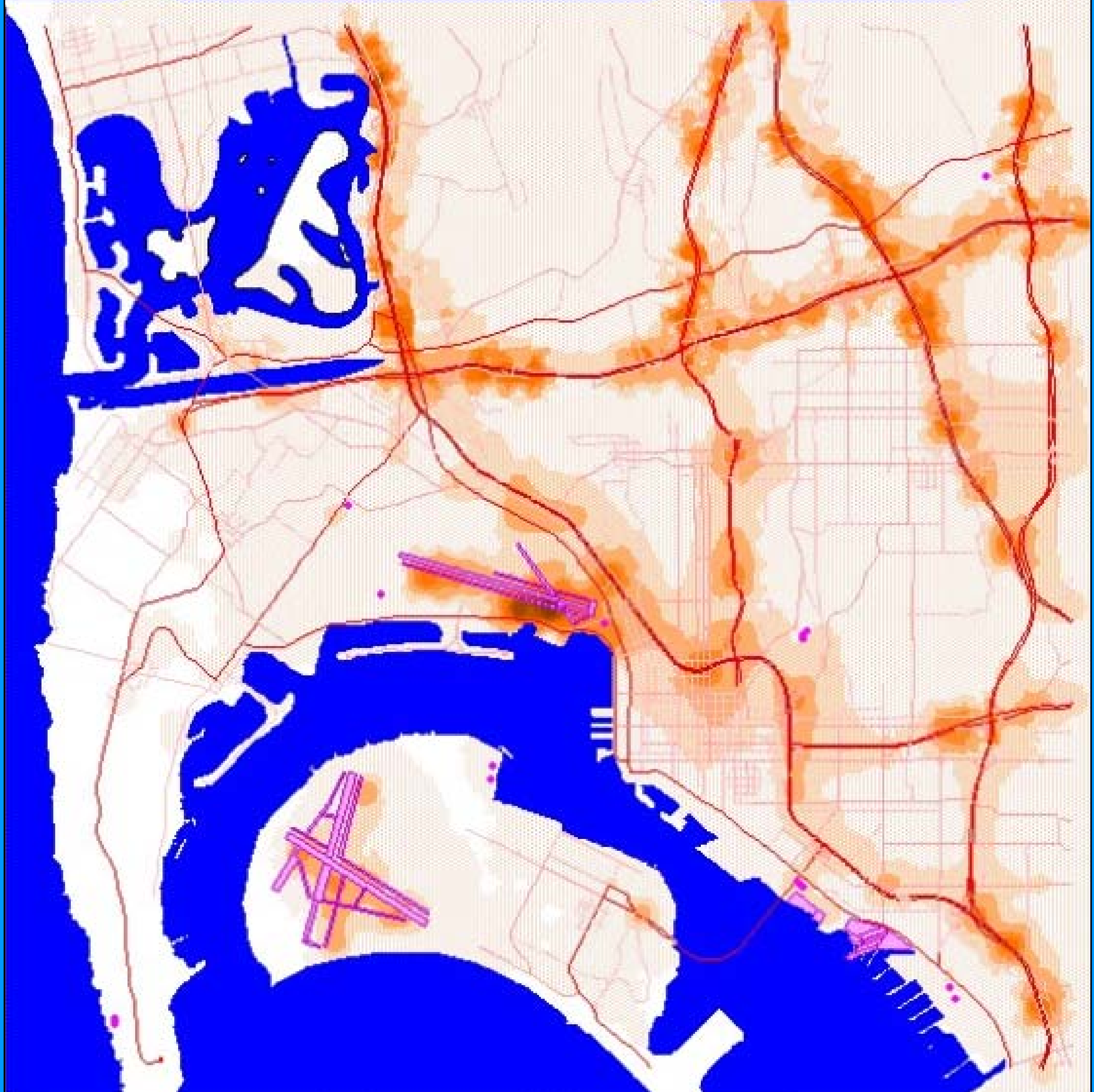
- Weaknesses:
  - Need enough monitoring locations for calibration/validation
  - Need to replicate in new areas; models developed in one geographic location may not be predictive in other areas.

Integrated Meterological-Emission Models (e.g. ADMS-Urban, Cal-PUFF)

– ADMS-Urban: Strengths:
  - Incorporates mobile, point, and area sources
  - Ability to model gridded emissions and terrain simultaneously
  - model 10,050 receptor points
  - boundary layer effects and dispersal behaviors over complex terrain
  - photochemistry.
  - seamlessly integrated with ArcView GIS

Integrated Meterological-Emission Models (e.g. ADMS-Urban, Cal-PUFF)

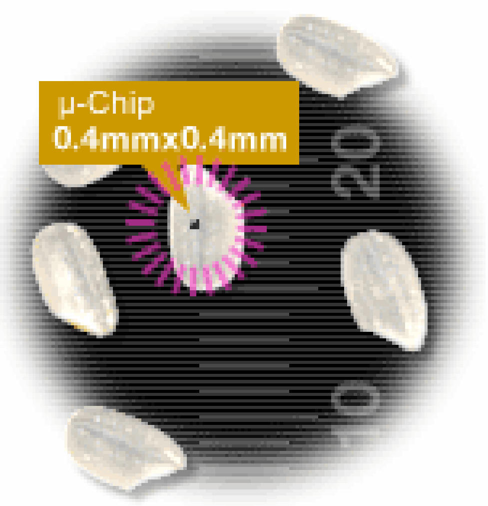
- ADMS-Urban:
  - Weaknesses::
    - Significant training and expertise necessary
    - Cost
    - Multiple data inputs
    - What is the bang for the buck?



Incorporating Time Activity/Personal Monitoring into Tracking

- Cost/sample size makes these activities prohibitive for tracking
  - Subsample analysis?
  - Survey data on time activities
  - Commuting important exposure time:
    - Use of Transportation demand models?

Specs:
128 bits ROM
Frequency: 2.45 GHz



*Size compared to a grain of rice

Conclusions

- Integrating density estimation techniques important for visualization and analysis of health tracking data – increased method development necessary
- Various approaches for traffic-exhaust modeling for tracking – Need to capture most accurate method over the most population at least cost