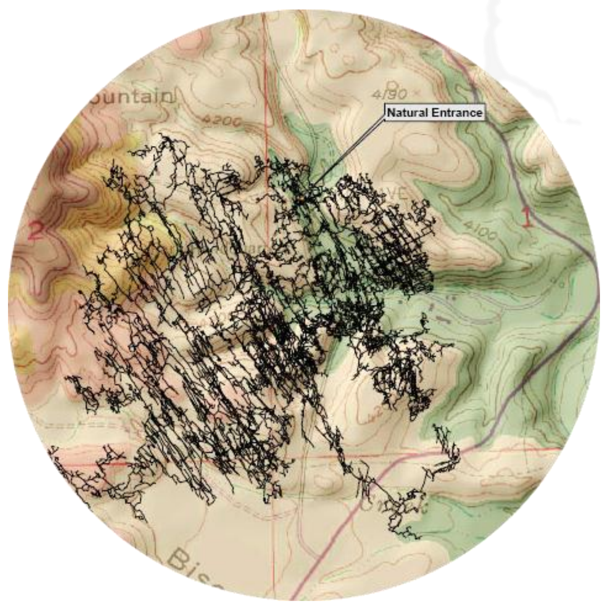
Analysis of

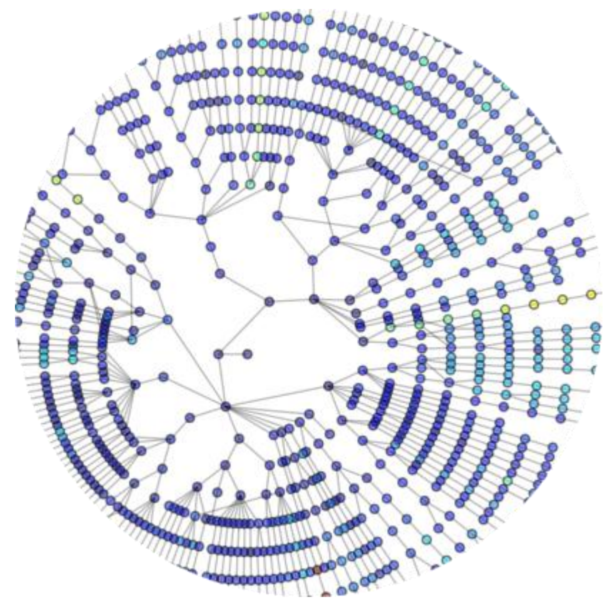
Geographically Embedded Networks

Ph.D. Dissertation Defense

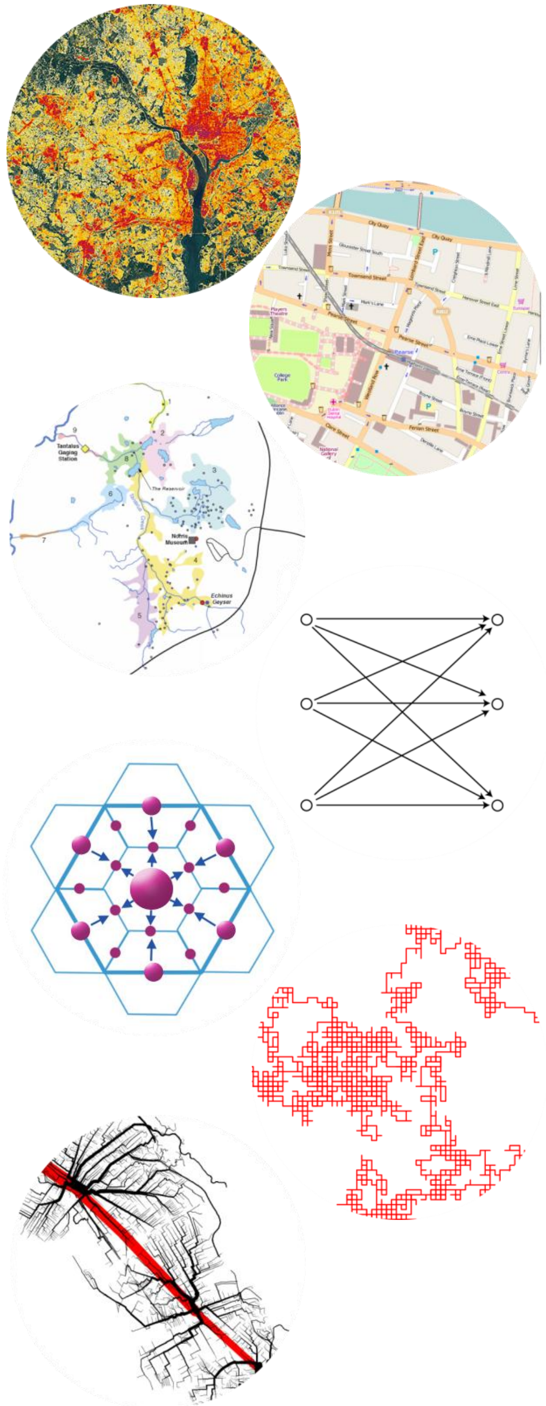
Alan Glennon
Department of Geography
University of California, Santa Barbara
11 January 2013

What are the distinguishing characteristics of networks in geography?





Kalinigrad, Russia by OpenStreetmap.org (2010)
Wind Cave by Bernie Szukalski and NPS (2010)
Lawrence Livermore IP routing graph by Graphserver (2010)



1 Introduction

2 Background

3 Data Modeling Use Cases

4 Analytical Use Cases

5 Properties of GENets

6 Use Case: Geyser Travel Problem

7 Contributions and Future Work

SLEUTH of Washington, D.C., by NASA & USGS (2004)
Dublin, Ireland, by OpenStreetMap.org (2010)
Tantalus Creek Hydrology by USGS (2007)
General Transportation Problem by Glennon
Christaller k=4 by Wikipedia.org (2010)
Random Walk by Wikipedia.org (2010)
Shortest Path by Graphserver project (2010)

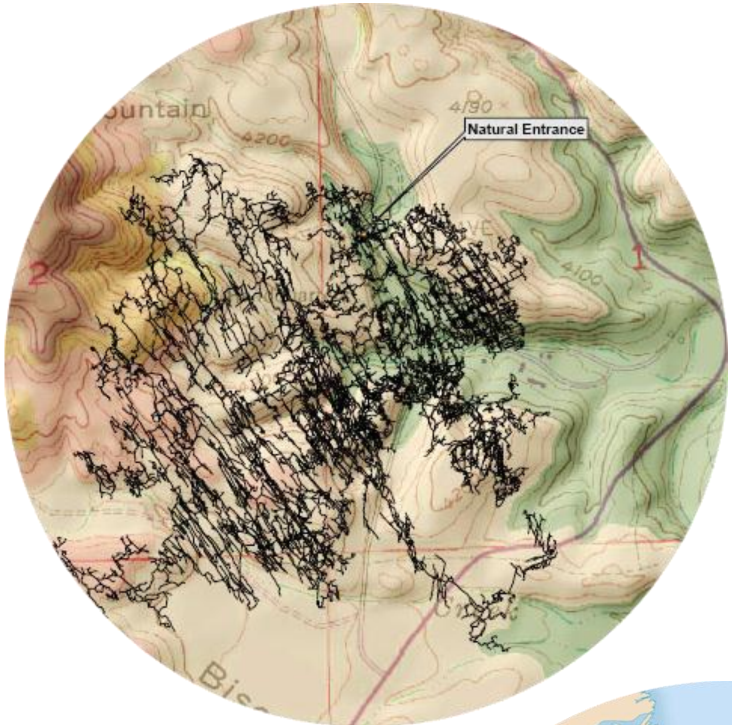
Geographically Embedded Network

“GENet”

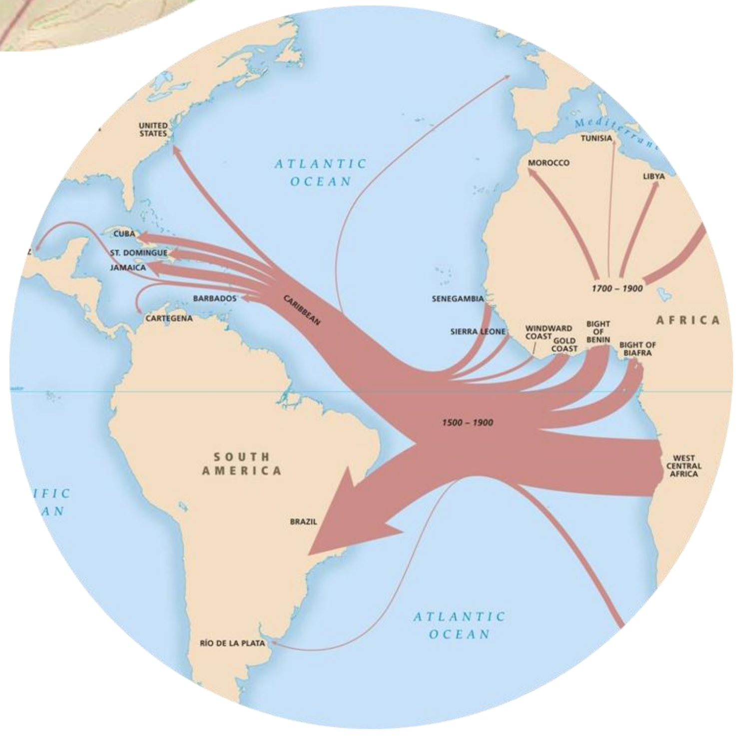
Image: Bonnie Long 2011; flickr.com/photos/68527152@N05/6235796437/

Types of GENets

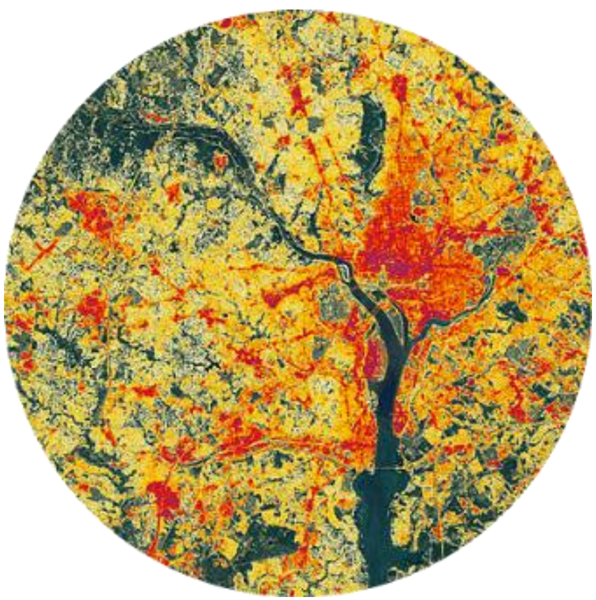
Physical GENet
Pathways tangibly exist in geographic space



Abstract GENet
Actors are embedded in geographic space, but pathways are conceptual



Source: Szukalski/NPS (2010) and Slavevoyages.org (2013)



Operations Research

Dijkstra, 1959
Moore, 1959
Ahuja, Magnanti, and Orlin, 1993

Transportation

Garrison 1958
Kansky, 1963
Haegerstrand, 1970

Hydrology

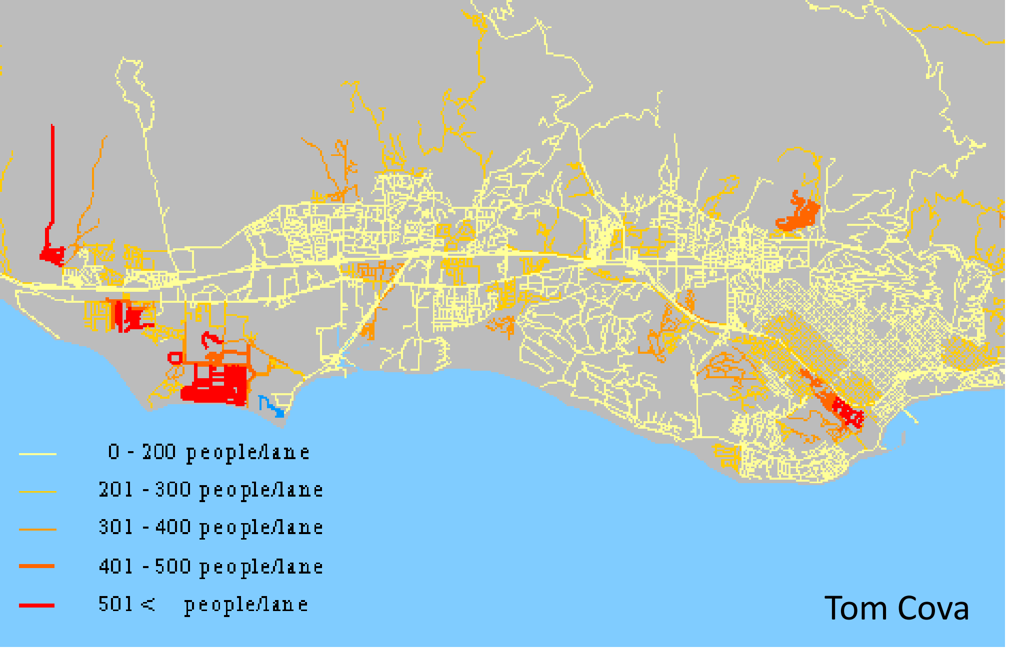
Horton, 1945
Shreve, 1967
Maidment, 2002

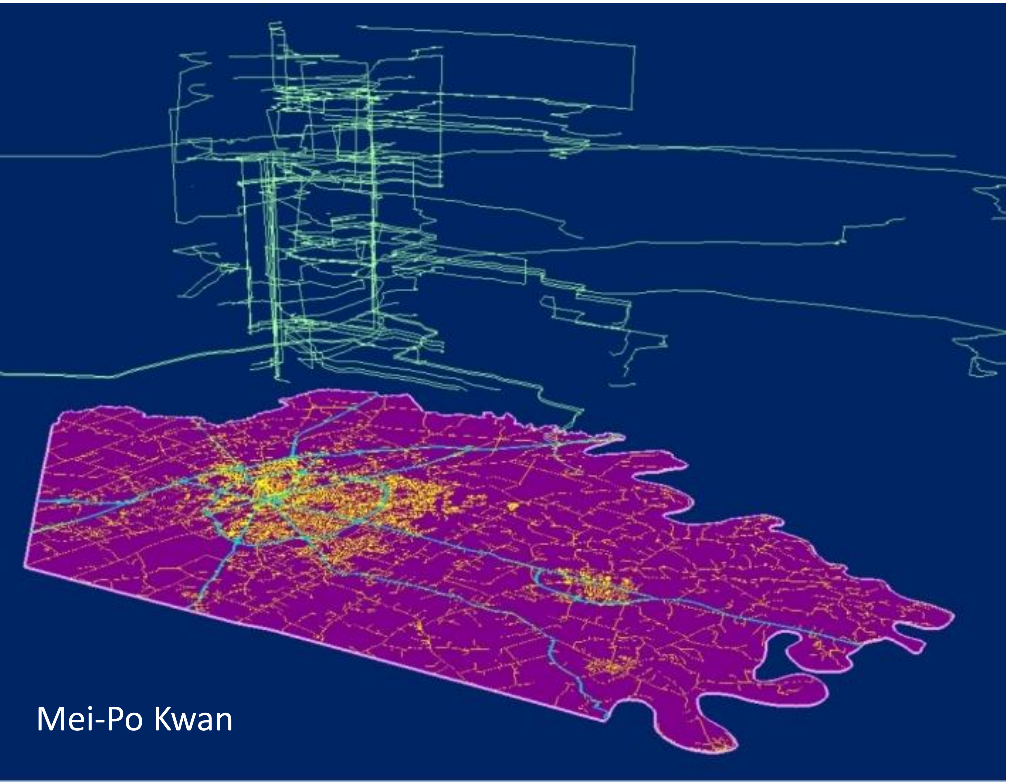
Simulation

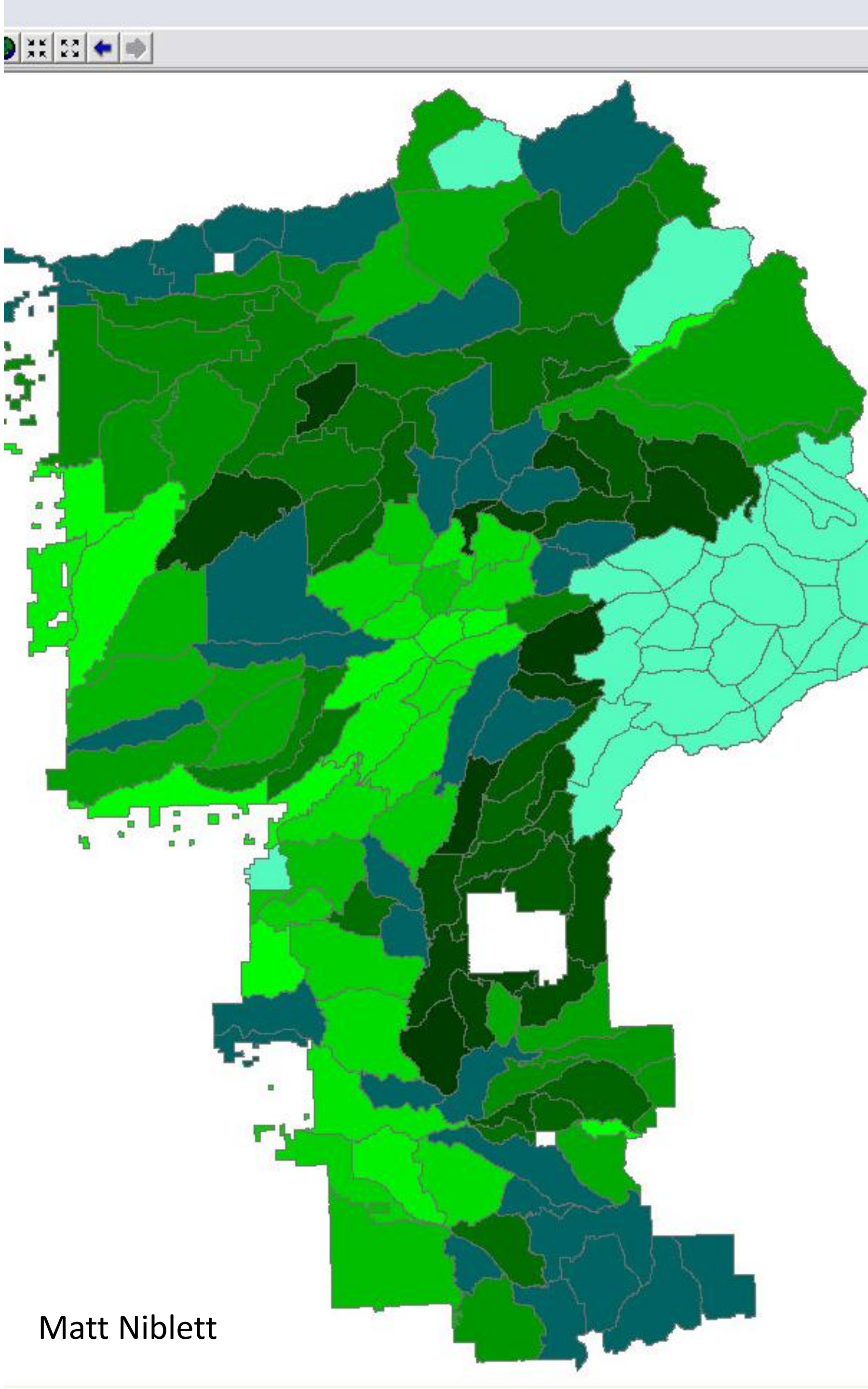
von Neuman, 1966
Conway, 1970
Wolfram, 2002

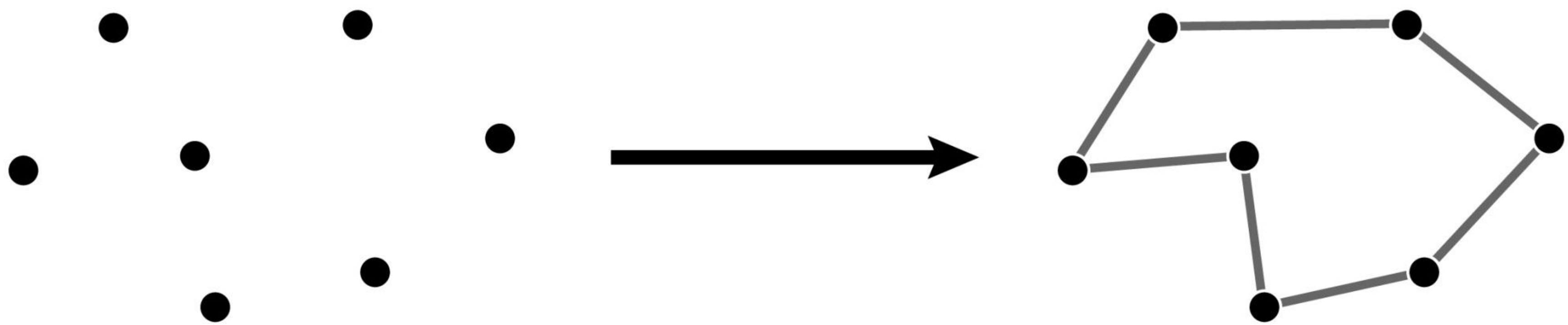
GIS

Tomlinson’s CGIS, 1970
Goodchild, 1991
Albrecht, 1997

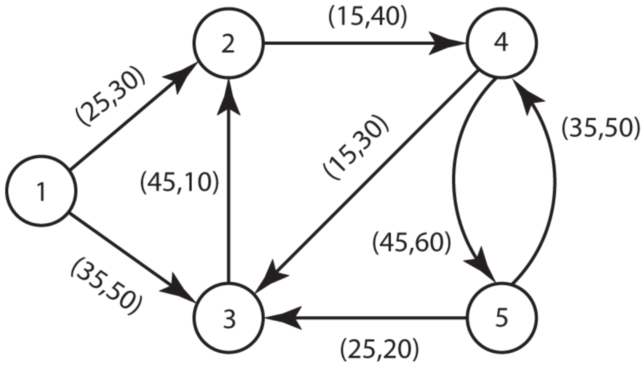

0 - 200 people/lane
201 - 300 people/lane
301 - 400 people/lane
401 - 500 people/lane
501 < people/lane
Tom Cova


Mei-Po Kwan


Matt Niblett



$G = (V, E)$



Reverse Star

	point
1	1
2	3
3	4
4	5
5	7
6	9

	tail	head	cost	capacity
1	1	2	25	30
2	1	3	35	50
3	2	4	15	40
4	3	2	45	10
5	4	3	15	30
6	4	5	45	60
7	5	3	25	20
8	5	4	35	50

Adjacency lists

Graph = { ’1’ : [’2’ , ’3’] , ’2’ : [‘4’] , ’3’ : [‘2’] , ‘4’ : [‘3’ , ‘4’] , ‘5’ : [‘3’ , ‘4’] }

Node-arc incidence matrix

$$
\begin{array}{c c}
 & \begin{array}{cccccccc} (1,2) & (1,3) & (2,4) & (3,2) & (4,3) & (4,5) & (5,3) & (5,4) \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\left[\begin{array}{cccccccc}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & -1 & 0 & -1 & 0 \\
0 & 0 & -1 & 0 & 1 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 & 1
\end{array}\right]
\end{array}
$$

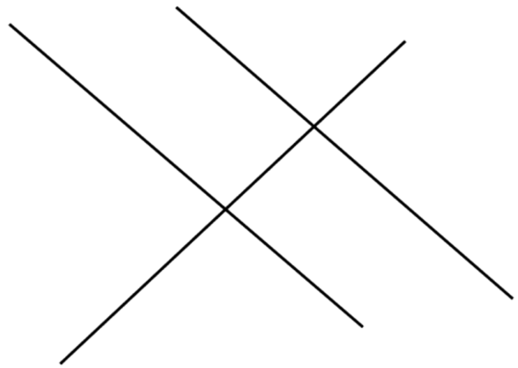
Node-node adjacency matrix

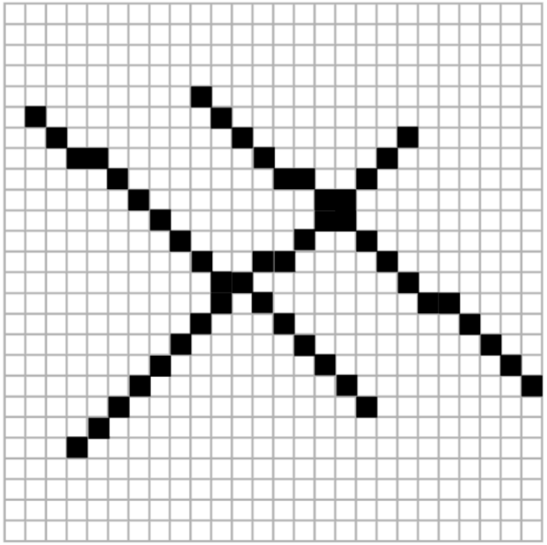
$$
\begin{array}{c c}
 & \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} &
\left[\begin{array}{ccccc}
0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 0
\end{array}\right]
\end{array}
$$

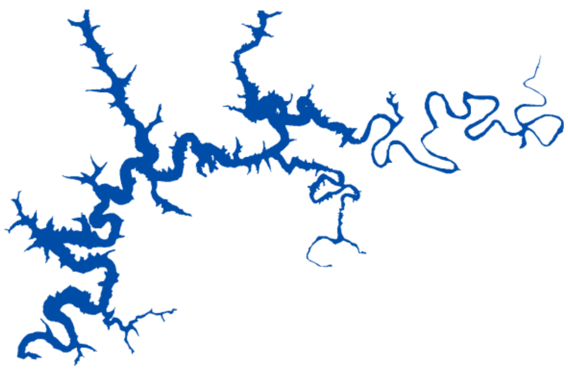
Computational Representations

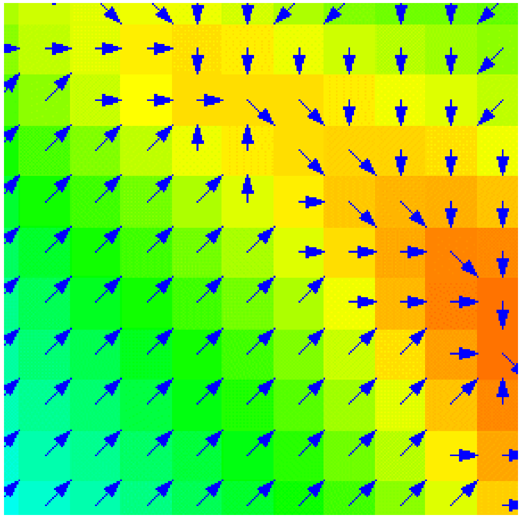
VECTOR

RASTER



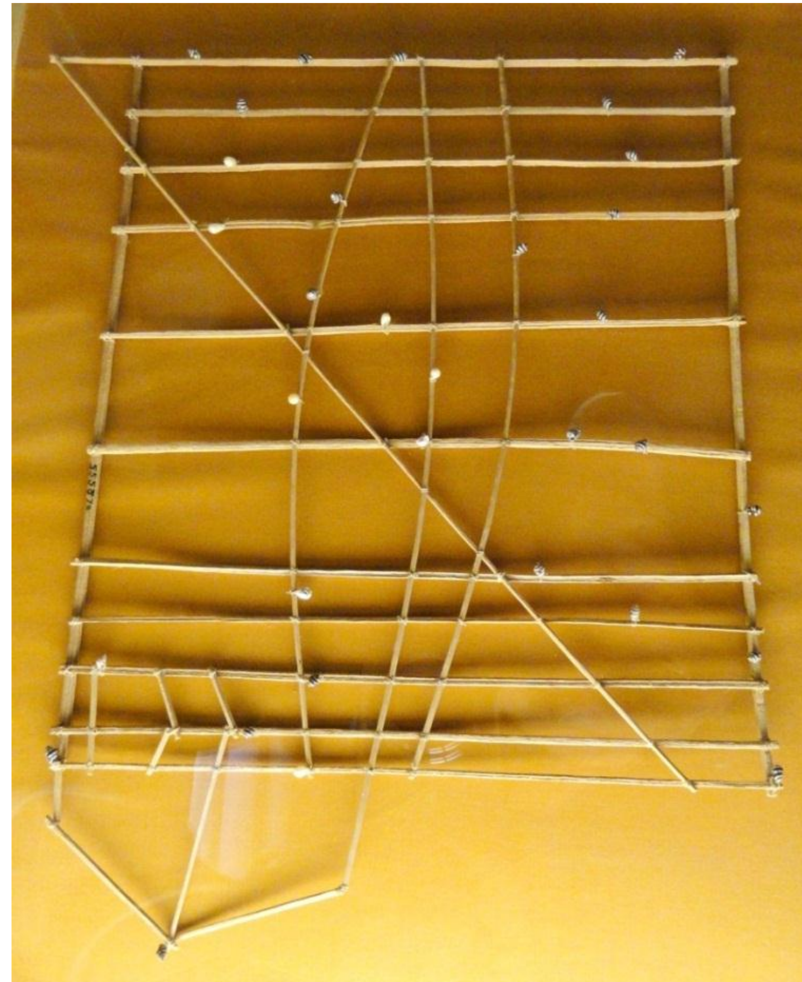




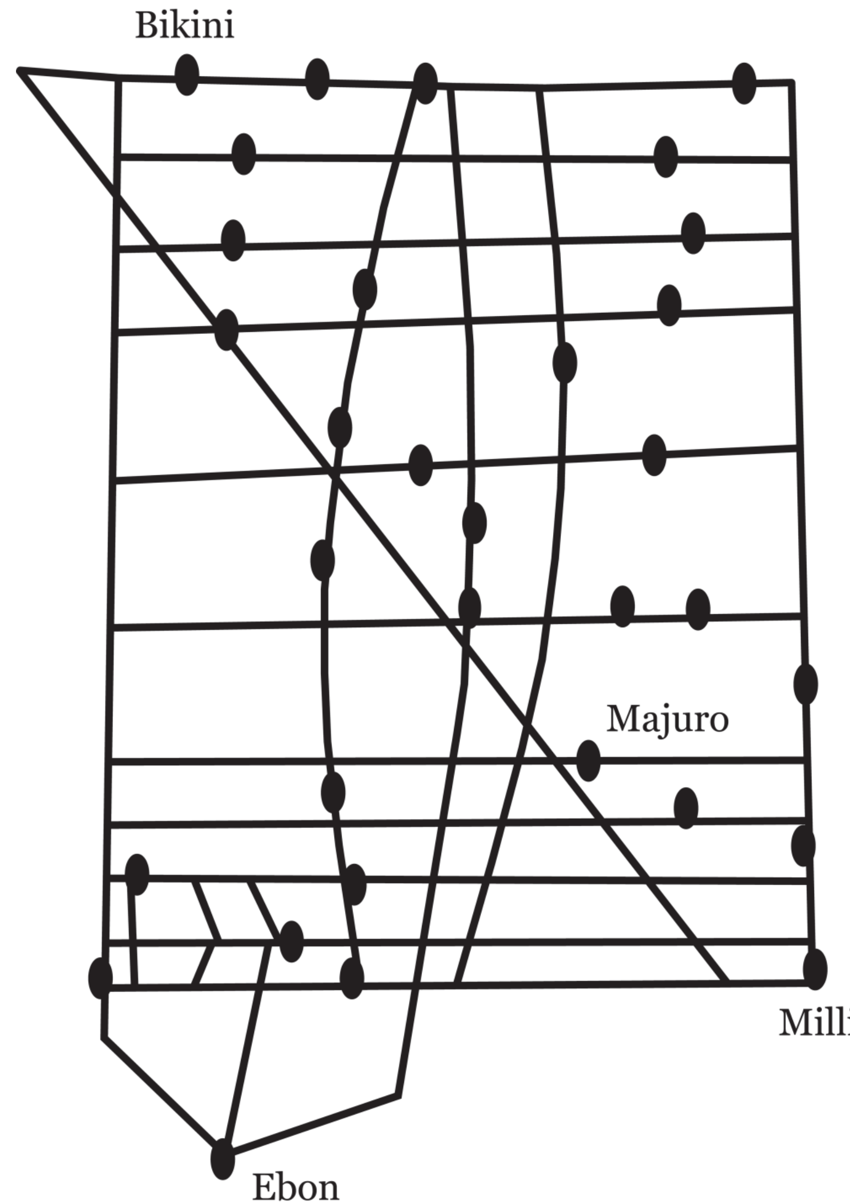


Source: http://www.flickr.com/photos/tofuart/7185212974/

Source: http://www.flickr.com/photos/tofuart/71852129747

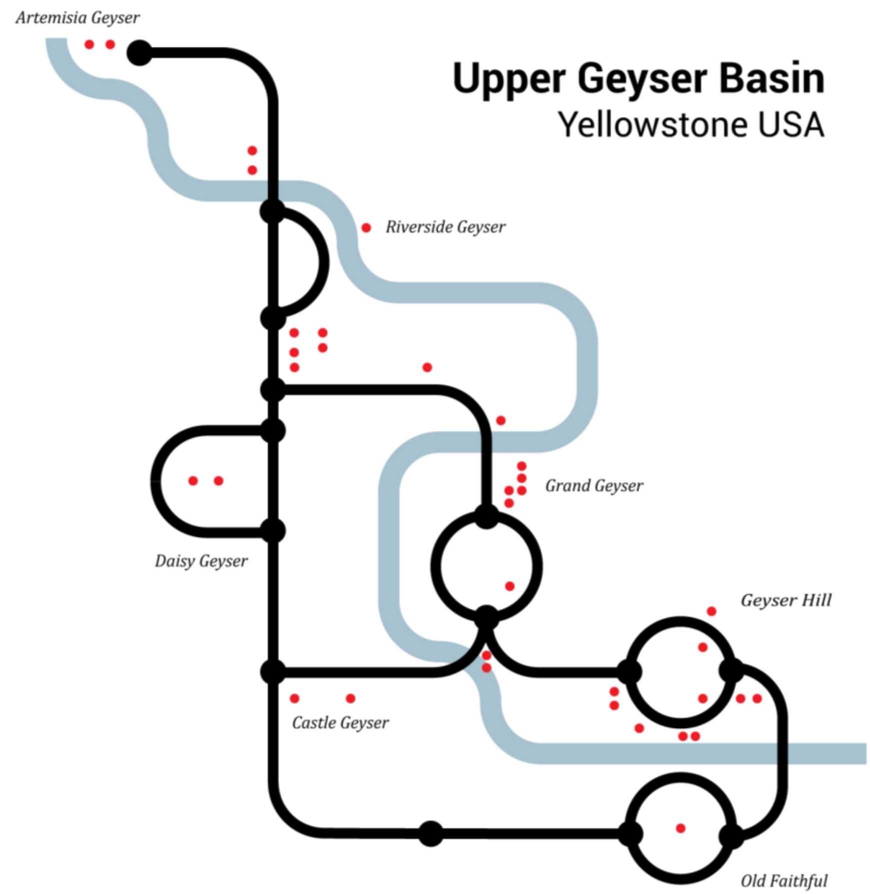


GRANTS CANAL
DELTA, LA.
VICKSBURG, MISS.



Bikini
Majuro
Milli
Ebon

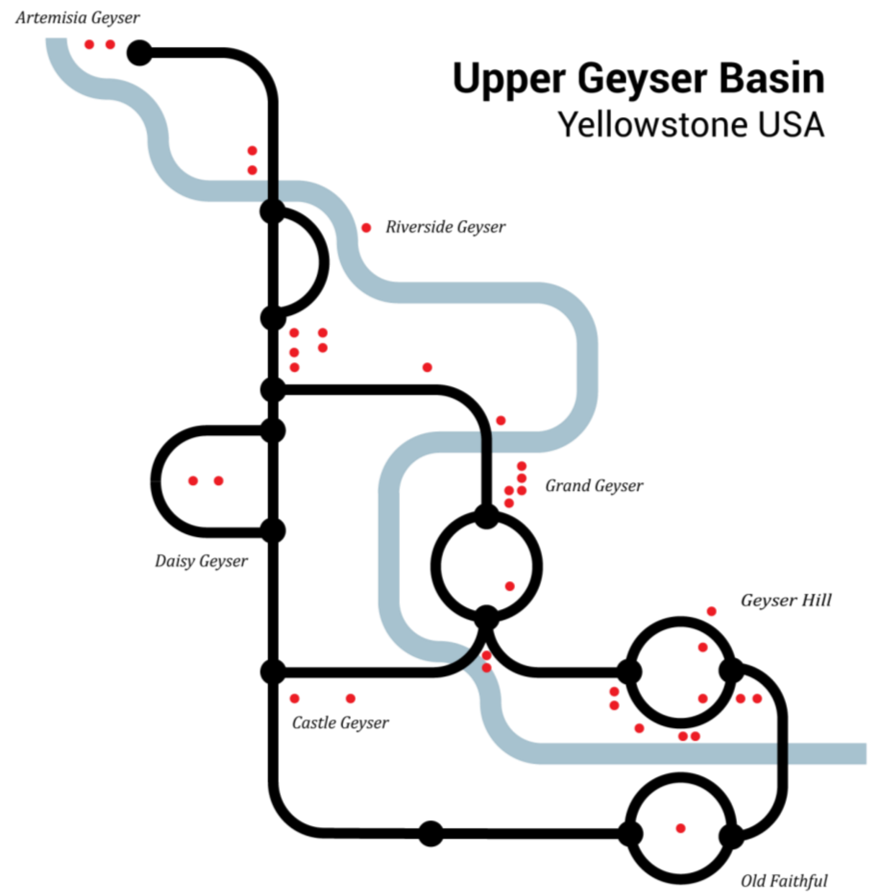
Edgware Road
Warren Street
Euston Square
Old Street
Regent's Park
Farringdon
Liverpool Street
Bayswater
Russell Square
Barbican
Goodge Street
Notting Hill Gate
Lancaster Gate
Bond Street
Oxford Circus
Chancery Lane
Moorgate
St. Paul's
Holland Park
Queensway
Marble Arch
Tottenham Court Road
Holborn
Bank
Covent Garden
Aldgate
High Street Kensington
Green Park
Leicester Square
Hyde Park Corner
Piccadilly Circus
Cannon Street
Monument
Tower Hill
Fenchurch Street
Knightsbridge
Mansion House
Charing Cross
Gloucester Road
Blackfriars
St. James's Park
Victoria
Temple
River Thames
Westminster
Embankment
London Bridge
Bermondsey
Earl's Court
South Kensington
Sloane Square
Waterloo
Southwark
Imperial Wharf
Pimlico
Lambeth North
Borough



Upper Geyser Basin
Yellowstone USA
Artemisia Geyser
Riverside Geyser
Grand Geyser
Daisy Geyser
Geyser Hill
Castle Geyser
Old Faithful

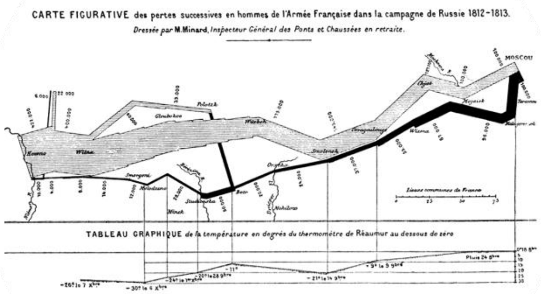

Edgware Road
Warren Street
Euston Square
Old Street
Regent's Park
Farringdon
Liverpool Street
Bayswater
Russell Square
Barbican
Notting Hill Gate
Lancaster Gate
Bond Street
Oxford Circus
Goodge Street
Chancery Lane
Moorgate
Holland Park
Queensway
Marble Arch
Tottenham Court Road
Holborn
St. Paul's
Bank
Aldgate
Covent Garden
High Street Kensington
Green Park
Leicester Square
Hyde Park Corner
Piccadilly Circus
Cannon Street
Monument
Tower Hill
Fenchurch Street
Knightsbridge
Mansion House
Charing Cross
Gloucester Road
Blackfriars
Victoria
St. James's Park
Temple
River Thames
Earl's Court
South Kensington
Sloane Square
Westminster
Embankment
London Bridge
Bermondsey
Waterloo
Southwark
Imperial Wharf
Pimlico
Lambeth North
Borough


CAULK CUT-OFF 1937
BEND
ISLAND NO. 77
ARKANSAS
MISSISSIPPI
JERSEY POINT
CATFISH POINT
MISSISSIPPI


Upper Geyser Basin
Yellowstone USA
Artemisia Geyser
Riverside Geyser
Grand Geyser
Daisy Geyser
Geyser Hill
Castle Geyser
Old Faithful

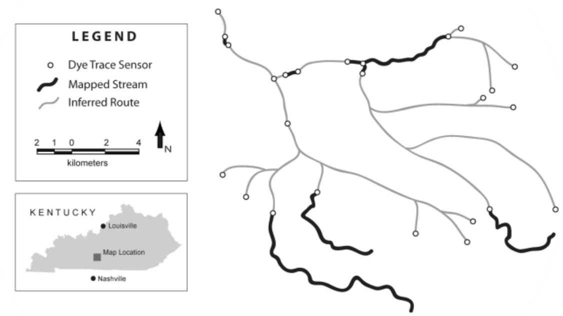
RIVER
ROSEDALE BEND
NAPOLEON CUT-OFF
ARKANSAS
DESHA
CO
MISS
BOLIVAR
COUNTY
LAKE BRULAH
MONTEREY BEND
INDIAN POINT
CAULK CUT-OFF 1937
CYPRESS BEND
CATFISH POINT
NECK
CAULK POINT
JERSEY POINT
BOLIVAR BEND
ARKANSAS
MISSISSIPPI
Ozark
Lobdell
Dahomy
Half Moon Bayou
Bolivar
Hannah

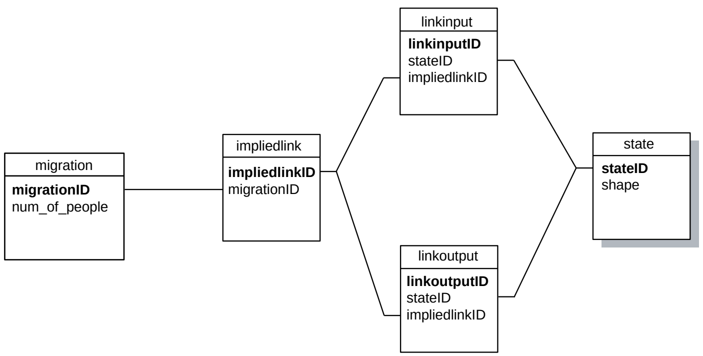
Data Modeling: GENet Flow

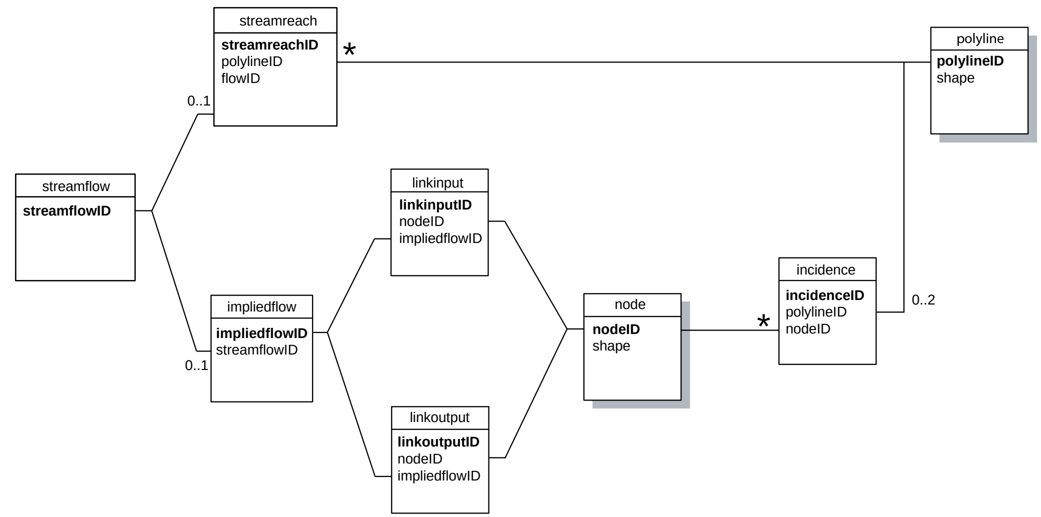


State of Origin	State of Destination	Migration Flow	Reverse Flow
New York	Florida	308,230	70,218
New York	New Jersey	206,979	97,584
California	Nevada	199,125	60,488
California	Arizona	186,151	92,452
California	Texas	182,789	115,929
Florida	Georgia	157,423	99,225
California	Washington	155,577	95,469
California	Oregon	131,836	67,642
New Jersey	Florida	118,905	34,896
Texas	California	115,929	182,789
New York	Pennsylvania	112,214	67,213
California	Colorado	111,322	56,050
New Jersey	Pennsylvania	110,436	88,202
New York	North Carolina	100,727	20,262
Georgia	Florida	99,225	157,423
New Jersey	New York	97,584	206,979
Florida	North Carolina	96,255	57,564
New York	California	95,952	65,160
Washington	California	95,469	155,577
California	Florida	94,265	65,211

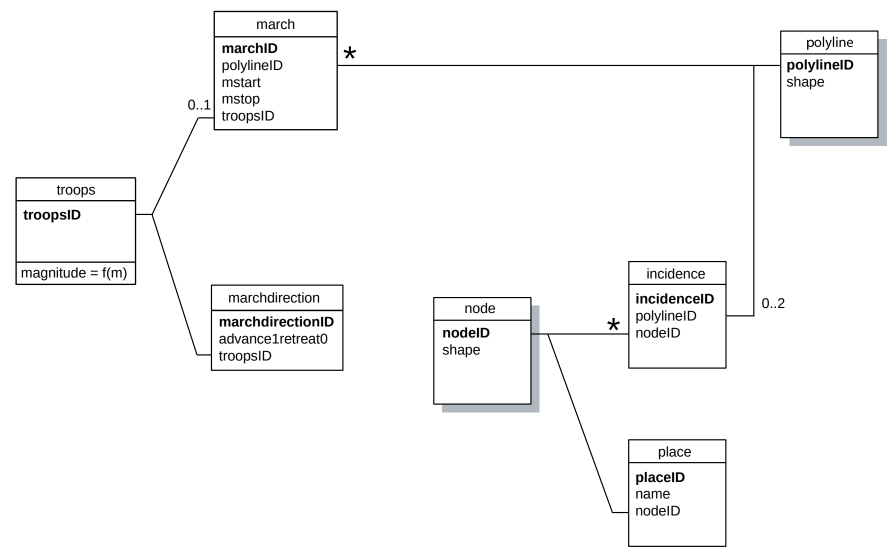
note: numbers are U.S. Census estimates from sample data.



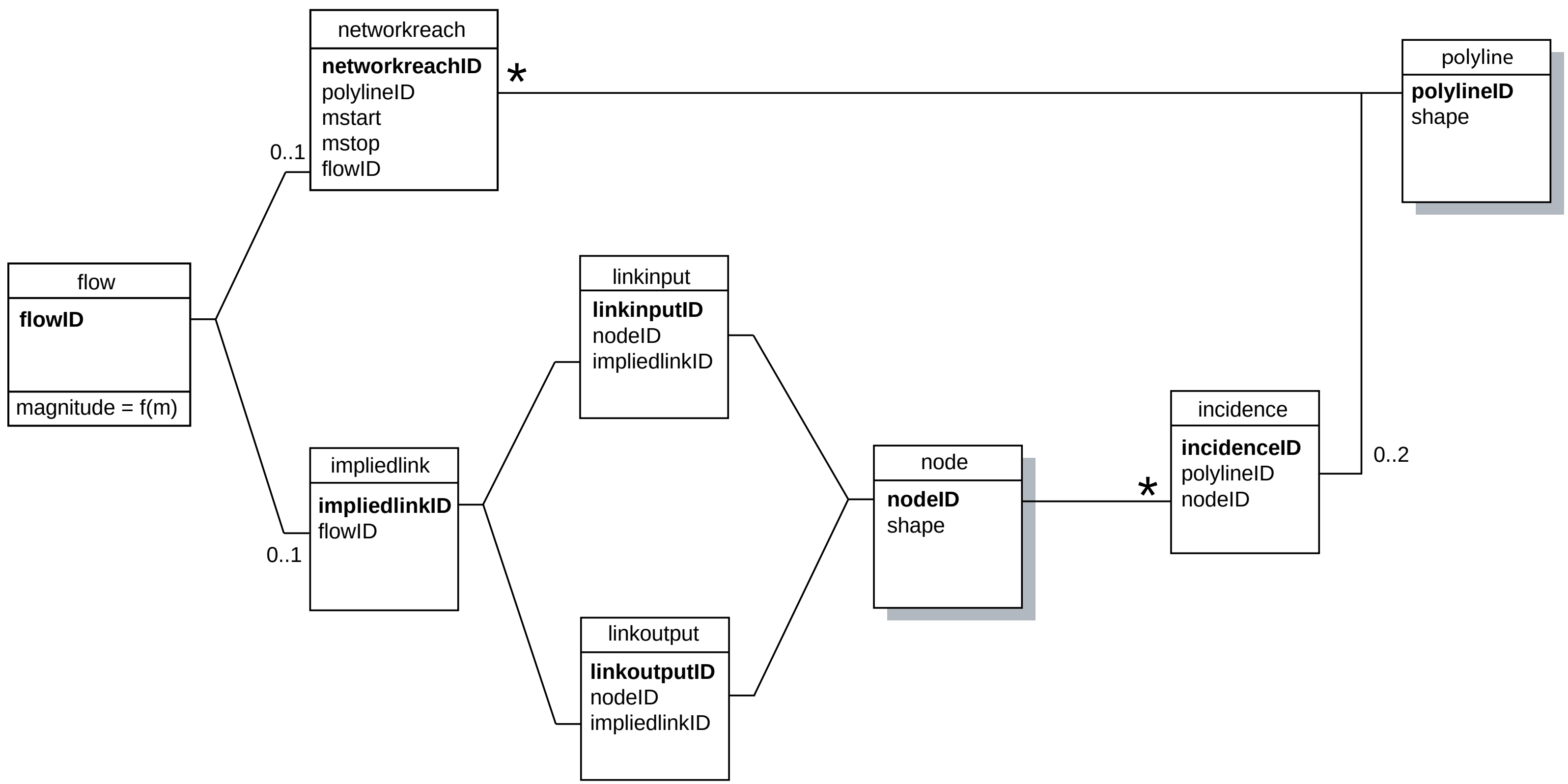


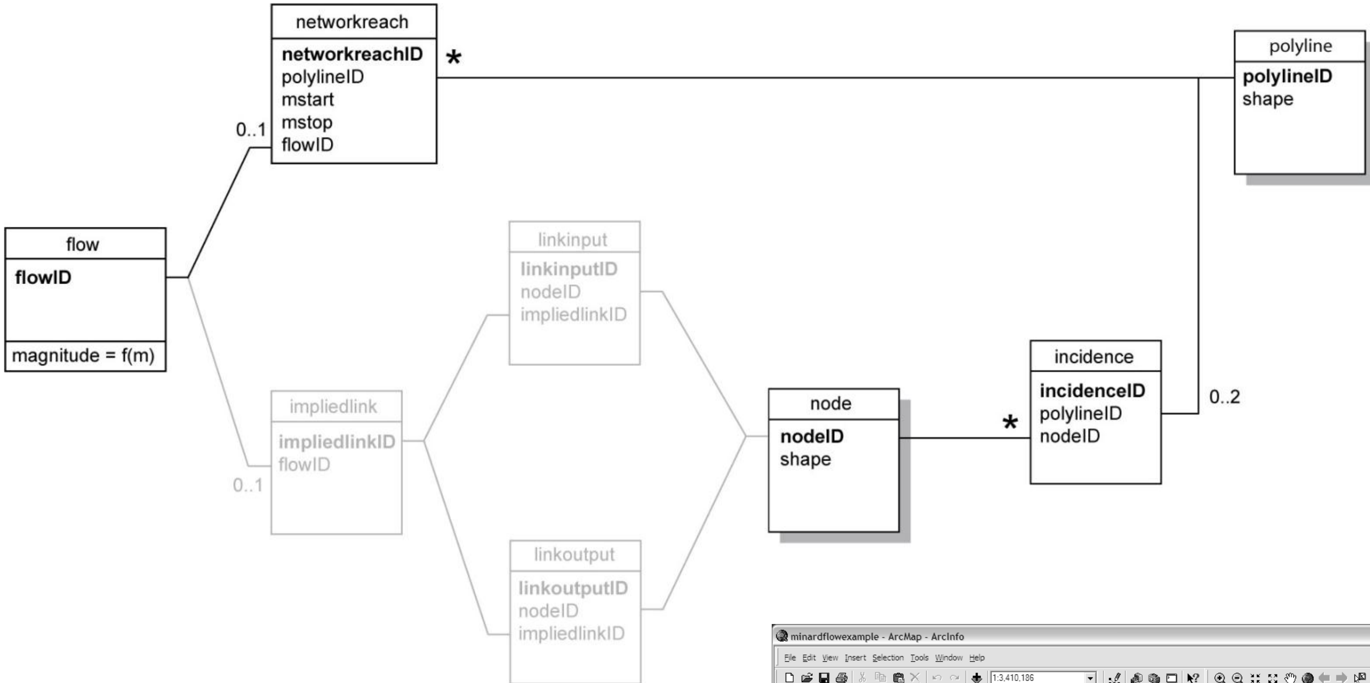


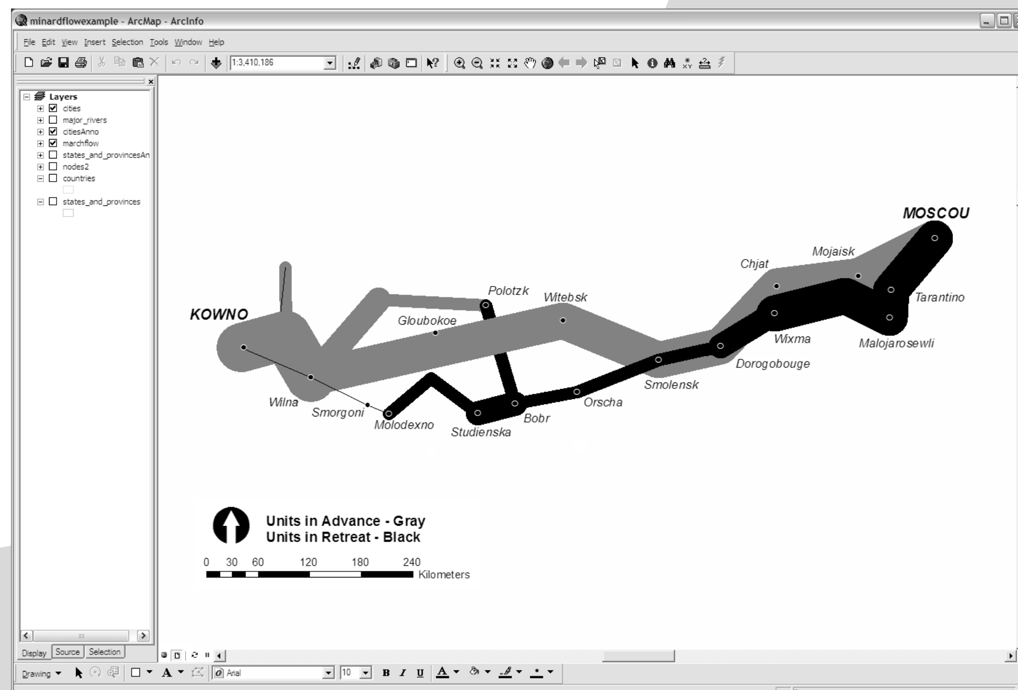
Use cases represented as UML data models

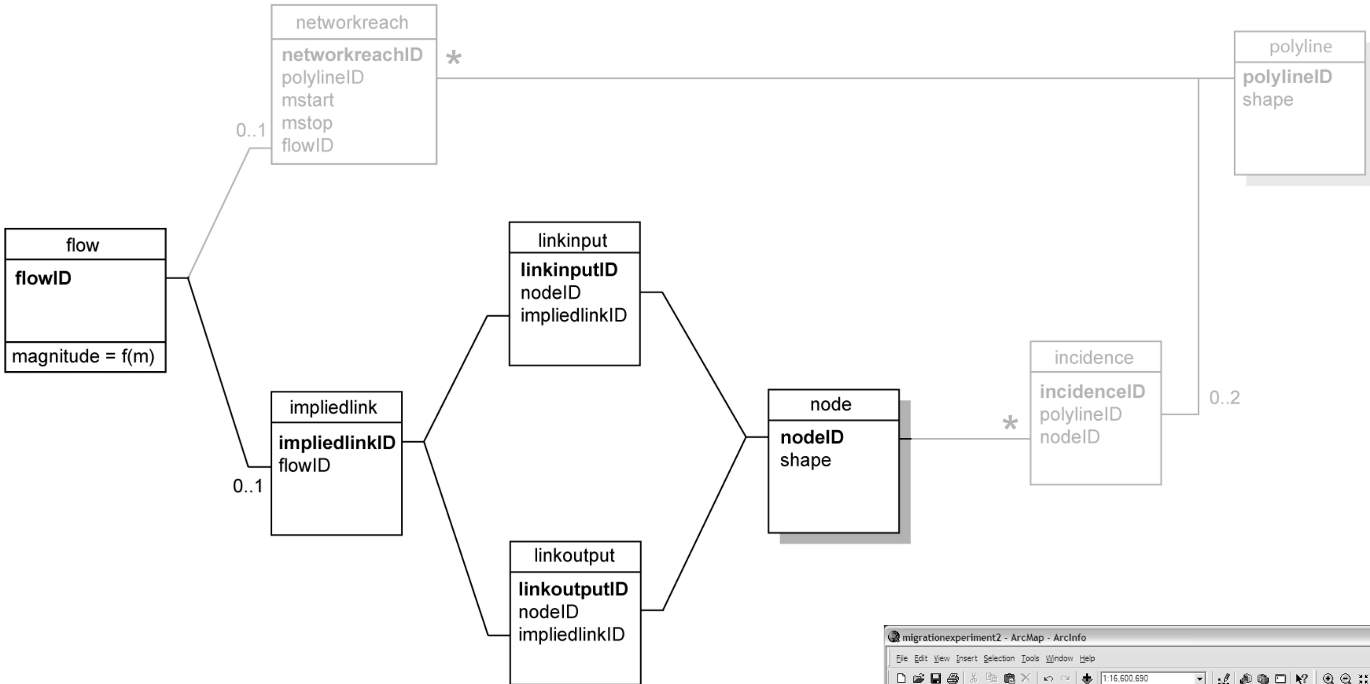


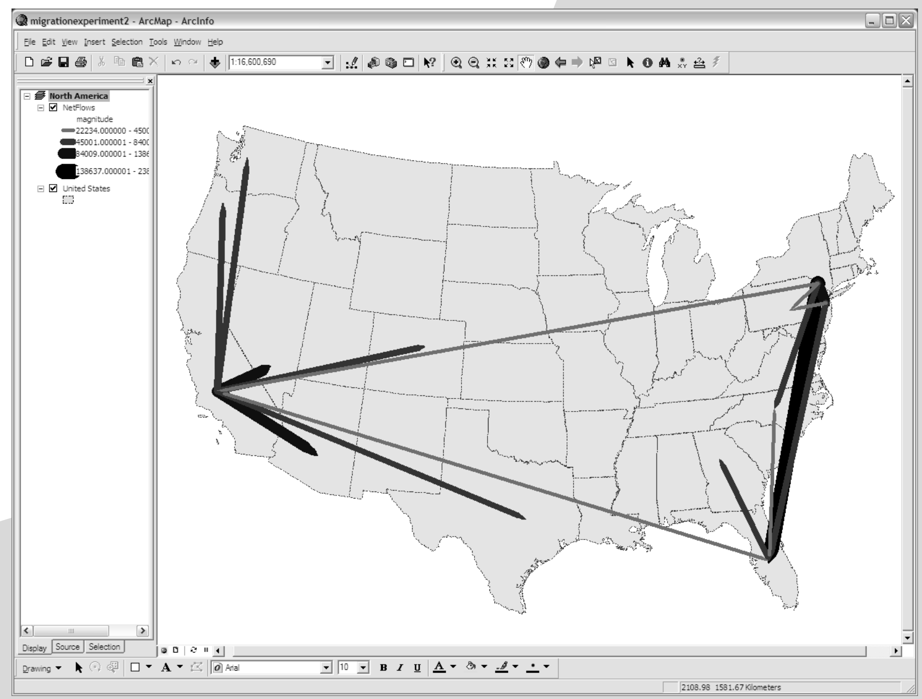
GENet Flow Data Model

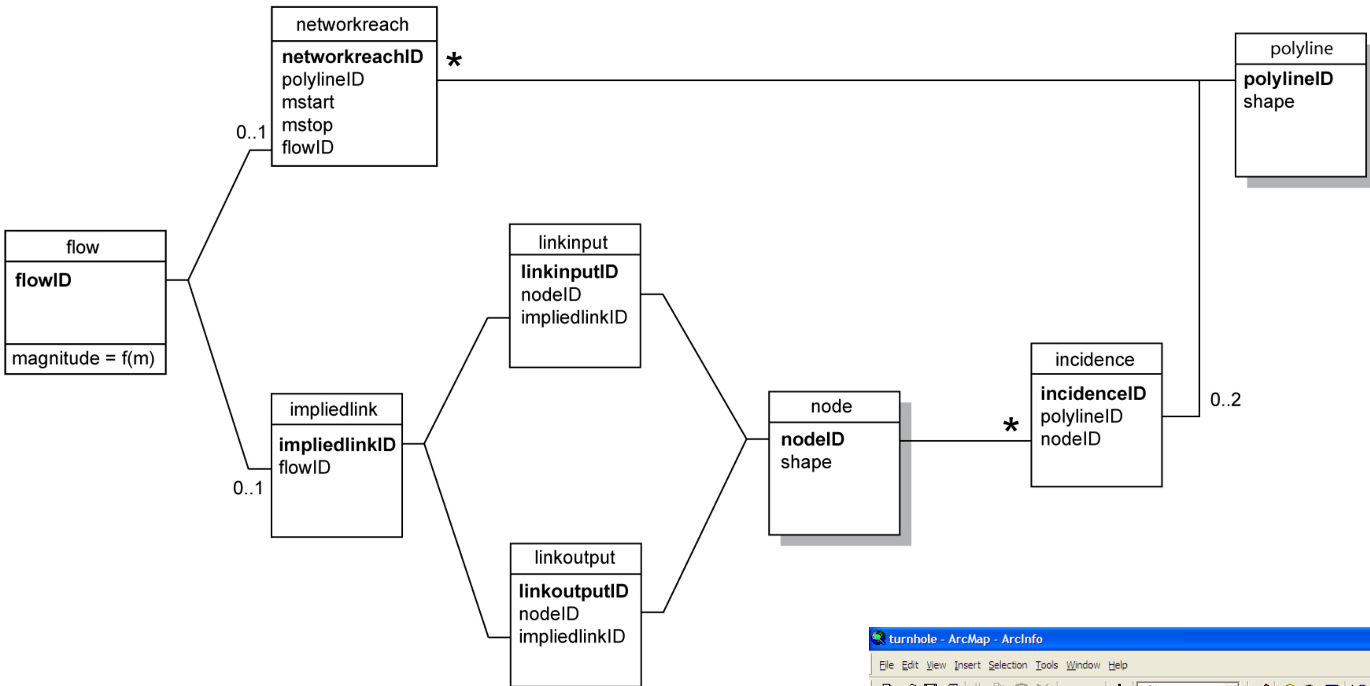


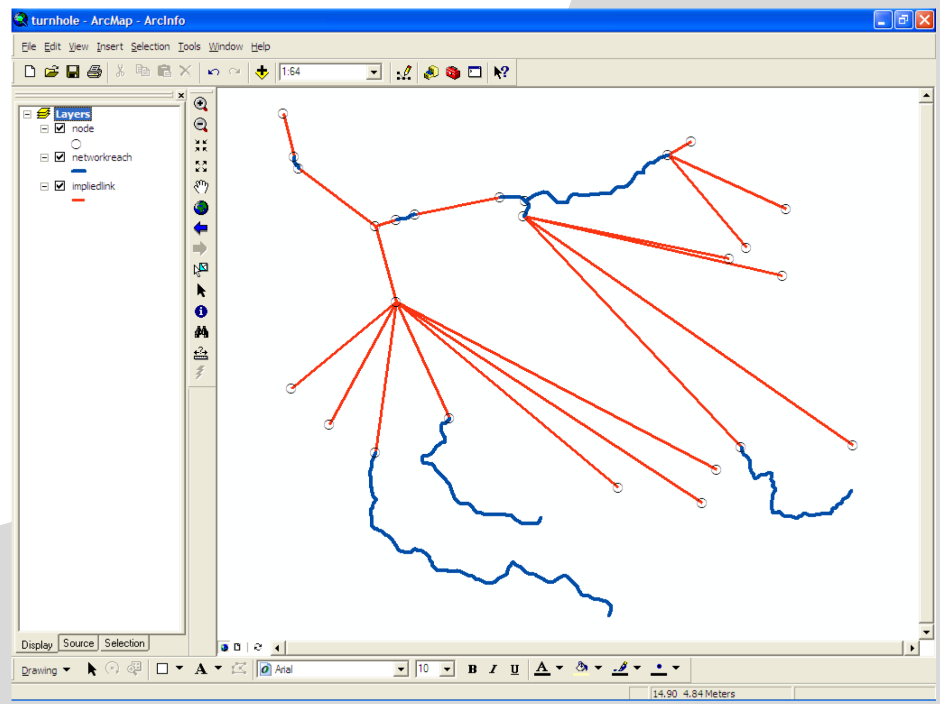

networkreach
networkreachID
polylineID
mstart
mstop
flowID
*
0..1
polyline
polylineID
shape
flow
flowID
magnitude = f(m)
linkinput
linkinputID
nodeID
impliedlinkID
impliedlink
impliedlinkID
flowID
0..1
incidence
incidenceID
polylineID
nodeID
0..2
node
nodeID
shape
*
linkoutput
linkoutputID
nodeID
impliedlinkID

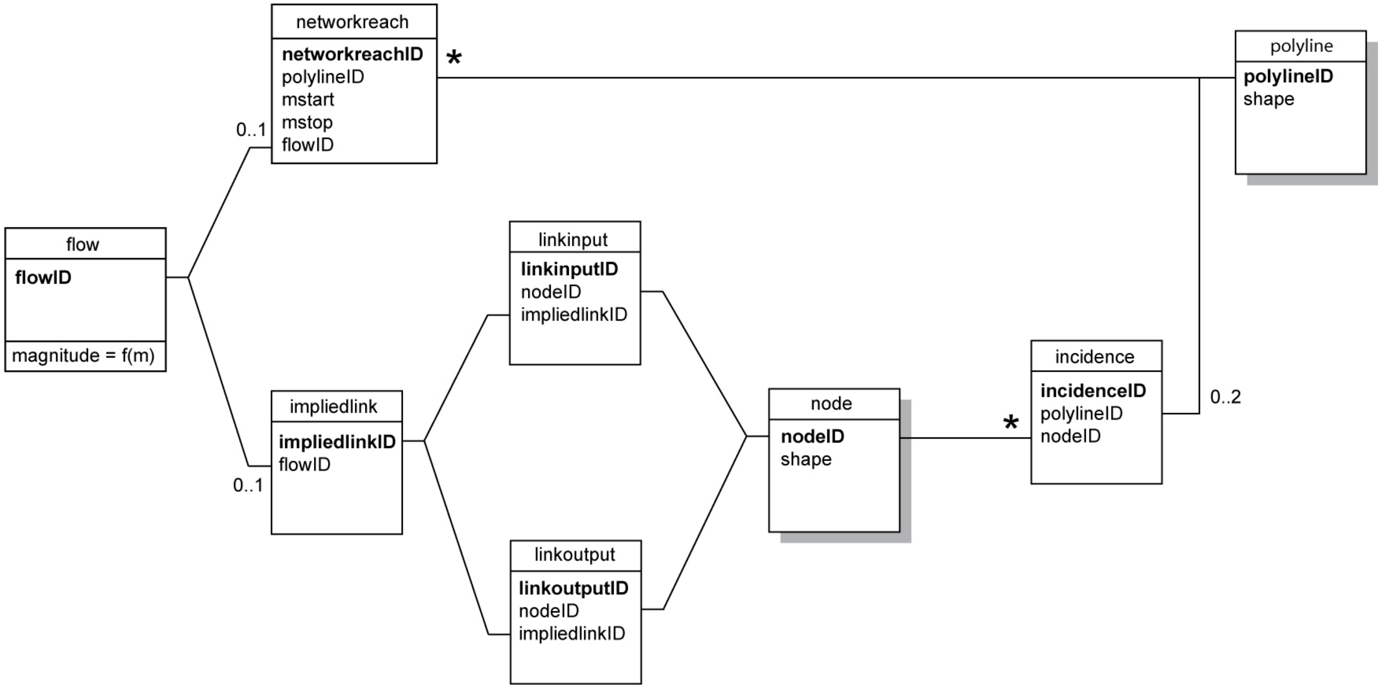

minardflowexample - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:3,410,186
Layers
cities
major_rivers
citiesAnno
marchflow
states_and_provincesAn
nodes2
countries
states_and_provinces
Display Source Selection
MOSCOU
Mojaisk
Chjat
Tarantino
Polotzk
Witebsk
KOWNO
Gloubokoe
Wixma
Malojarosewli
Dorogobouge
Smolensk
Wilna
Smorgoni
Molodexno
Studienska
Bobr
Orscha
Units in Advance - Gray
Units in Retreat - Black
0 30 60 120 180 240
Kilometers
Drawing
Arial
10


networkreach
networkreachID
polylineID
mstart
mstop
flowID
*
0..1
polyline
polylineID
shape
flow
flowID
magnitude = f(m)
linkinput
linkinputID
nodeID
impliedlinkID
impliedlink
impliedlinkID
flowID
0..1
node
nodeID
shape
incidence
incidenceID
polylineID
nodeID
*
0..2
linkoutput
linkoutputID
nodeID
impliedlinkID


migrationexperiment2 - ArcMap - ArcInfo
North America
NetFlows
magnitude
United States


networkreach
networkreachID
polylineID
mstart
mstop
flowID
*
polyline
polylineID
shape
0..1
flow
flowID
magnitude = f(m)
linkinput
linkinputID
nodeID
impliedlinkID
incidence
incidenceID
polylineID
nodeID
0..2
impliedlink
impliedlinkID
flowID
0..1
node
nodeID
shape
*
linkoutput
linkoutputID
nodeID
impliedlinkID

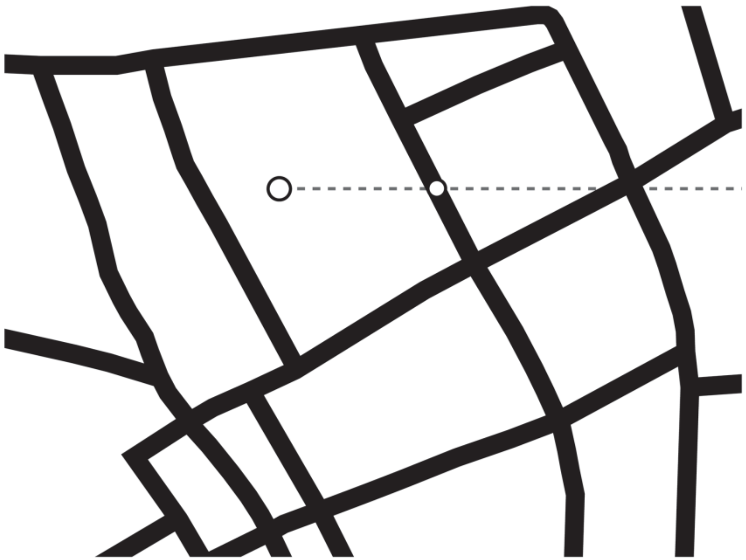

turnhole - ArcMap - ArcInfo
File Edit View Insert Selection Tools Window Help
1:64
Layers
node
networkreach
impliedlink
Display
Source
Selection
Drawing
Arial
10
14.90 4.84 Meters

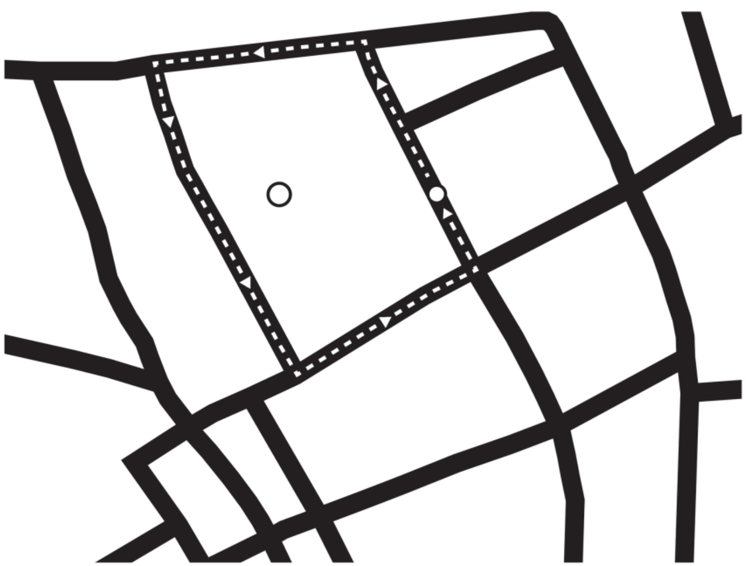


Mapped and uncertain routes

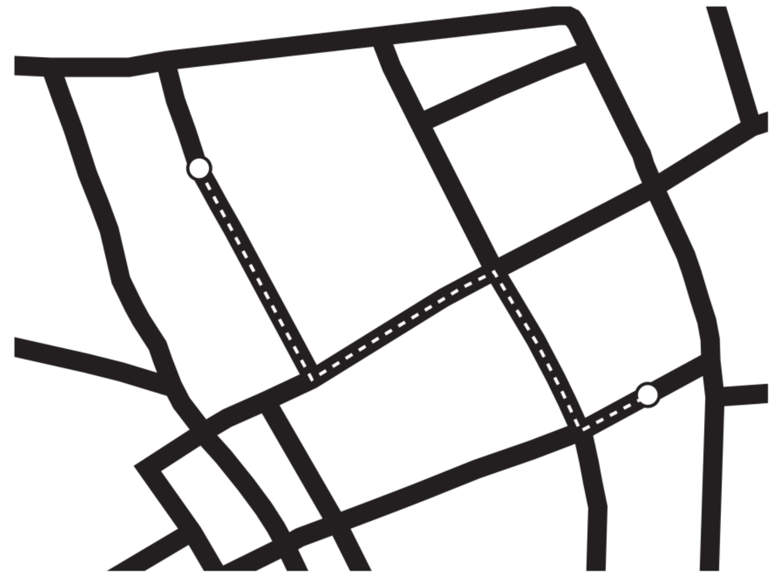
Dynamic attributes as a function of space

Selecting streets bounding a city block





Calculating shortest path



NETWORK ANALYSIS

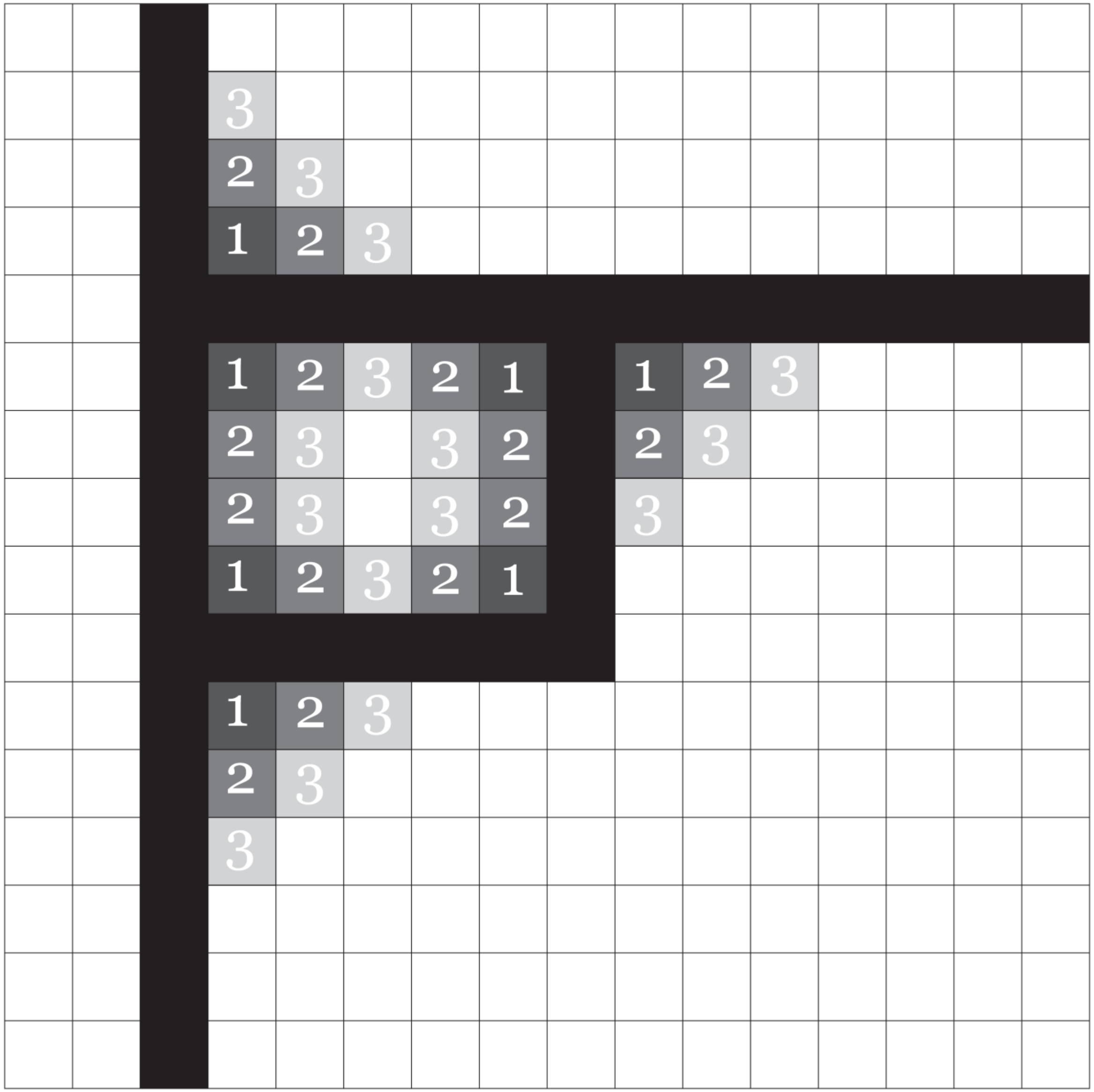
Comparison

Optimization

Simulation

Process

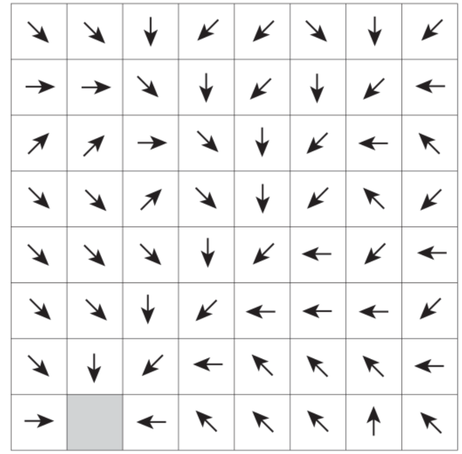
Simulating urban growth near a road



Constructing a stream network from a Digital Elevation Model

10	9	8	9	12	11	12	13
9	8	7	8	10	9	8	12
9	9	8	6	8	7	10	11
10	9	8	7	5	8	9	10
9	8	7	7	4	8	8	11
8	7	5	3	5	6	9	9
4	3	2	4	6	9	7	9
2	1	3	6	7	8	9	10

(a)



(b)

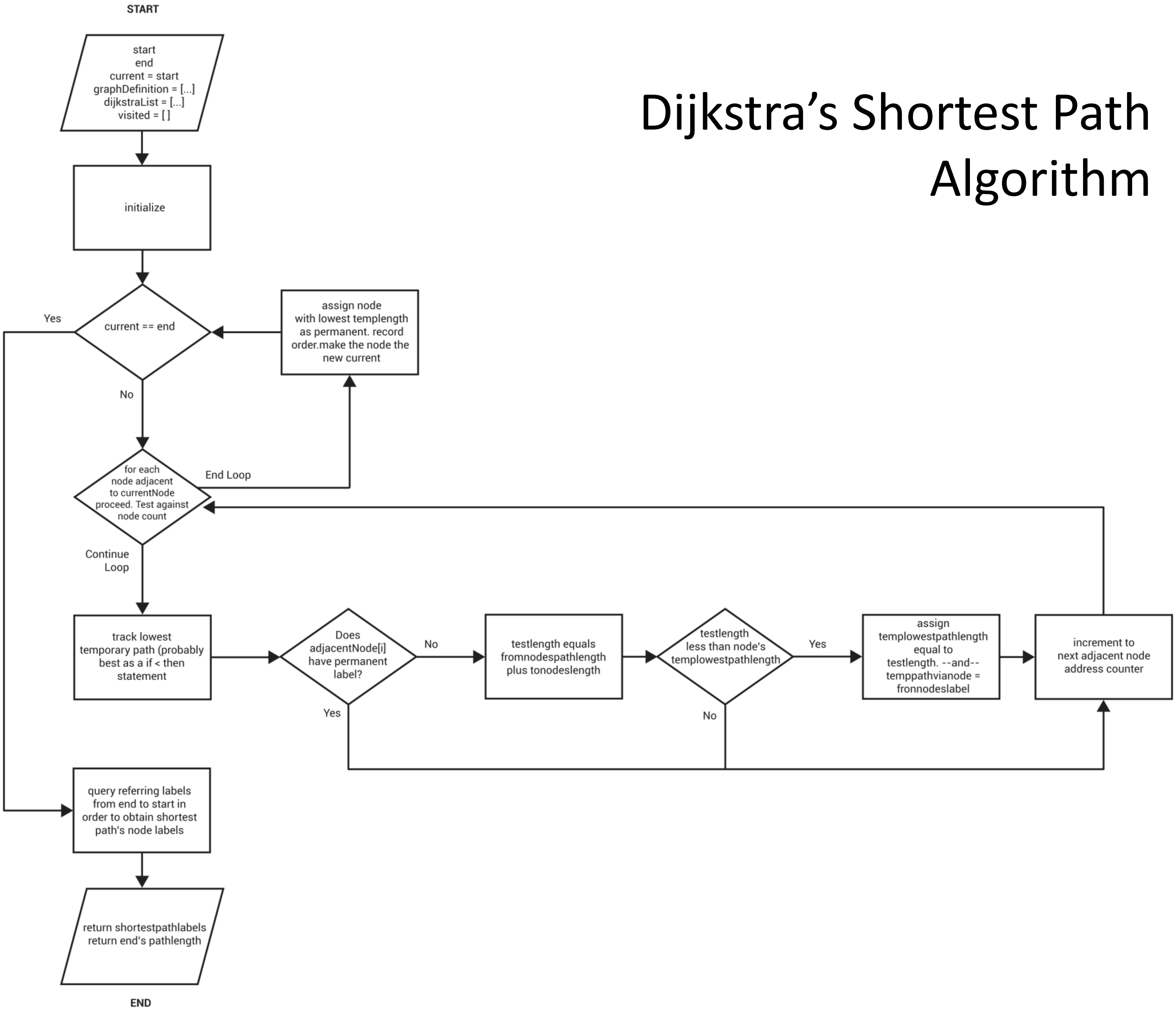
0	0	0	0	0	0	0	0
0	3	8	1	0	0	5	0
0	0	0	14	0	9	0	0
0	0	0	0	26	0	0	0
0	1	1	0	30	0	2	0
0	1	2	48	11	9	0	0
0	1	57	1	1	0	4	0
0	63	0	0	0	0	0	0

(c)

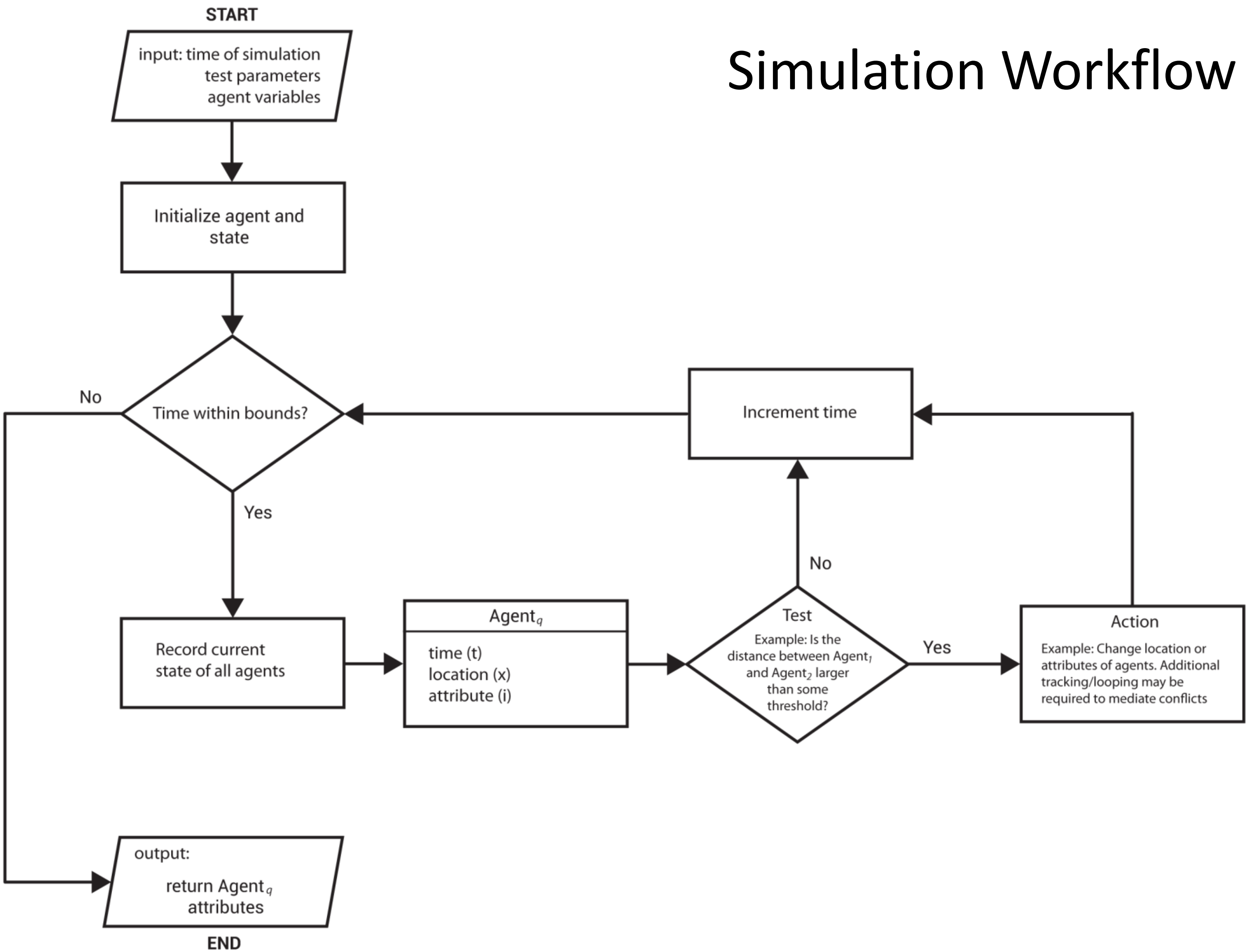
0	0	0	0	0	0	0	0
0	3	8	1	0	0	5	0
0	0	0	14	0	9	0	0
0	0	0	0	26	0	0	0
0	1	1	0	30	0	2	0
0	1	2	48	11	9	0	0
0	1	57	1	1	0	4	0
0	63	0	0	0	0	0	0

(d)

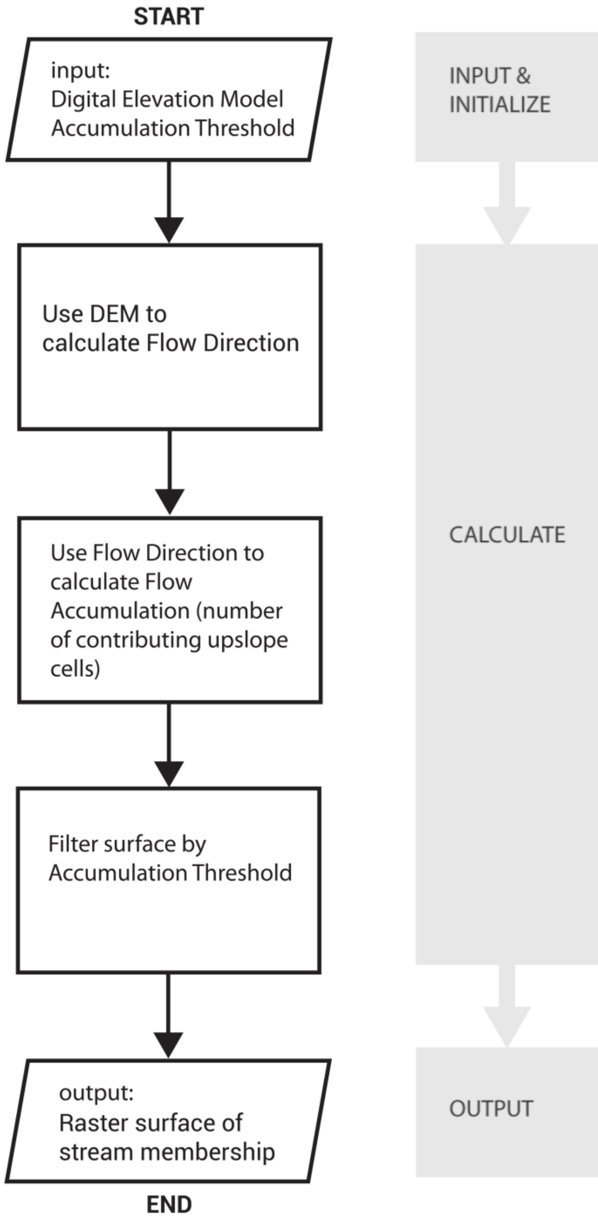
Dijkstra’s Shortest Path Algorithm



Simulation Workflow



Workflow for constructing a stream network from a Digital Elevation Model



Parameters
For characterizing
GENet operations
(modified from Mitchell 2000)

Objective

What is the problem objective?
What are the problem constraints?

Data

How is the network represented?
Does the data support uncertain, fuzzy, or missing data?

Algorithm

Are operational rules deterministic or stochastic?
Does the algorithm require iteration?
    If so, what is its nature (finite, continuous, dynamic feedback)?
What dimensions does the algorithm consider?
Is the environment static or dynamic?
Does the algorithm use exact or approximate methods?
How does the algorithm handle uncertain or missing data?

Output

What is the nature of the solution? For instance, is it vector, raster, numeric, descriptive?
Is the solution a subset of existing data or something newly derived?

Process

Is scale a consideration? For instance, does the answer change with varying scale?
Is the algorithm designed for single or repeated use?
Are the results repeatable?
Is the process reversible without data loss?
How are distance units or length addressed?

Is a model spatial? four possible tests (Goodchild 2012)

invariance test
spatially explicit models are variant under relocation of the objects of study

representation test
spatially explicit models include representation of location in their implementations

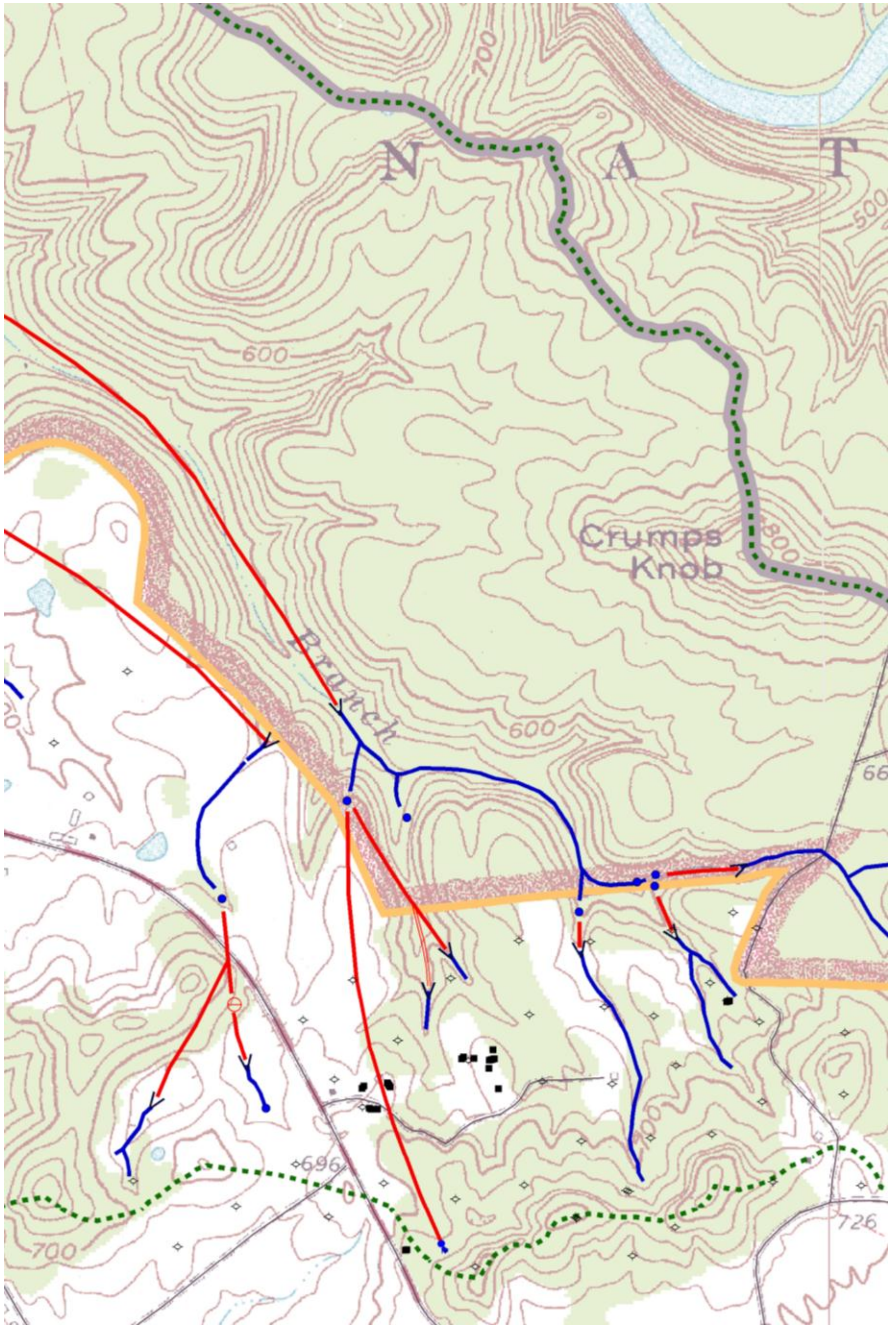
formulation test
concepts such as location or distance appear directly in the model in algebraic expressions or behavioral rules

outcome test
spatial structures of inputs and outputs are different
it modifies the landscape on which it operates

Properties of GENets

REPRESENTATIONAL

PHYSICAL



Source: USGS and Glennon

Properties of GENets

REPRESENTATIONAL

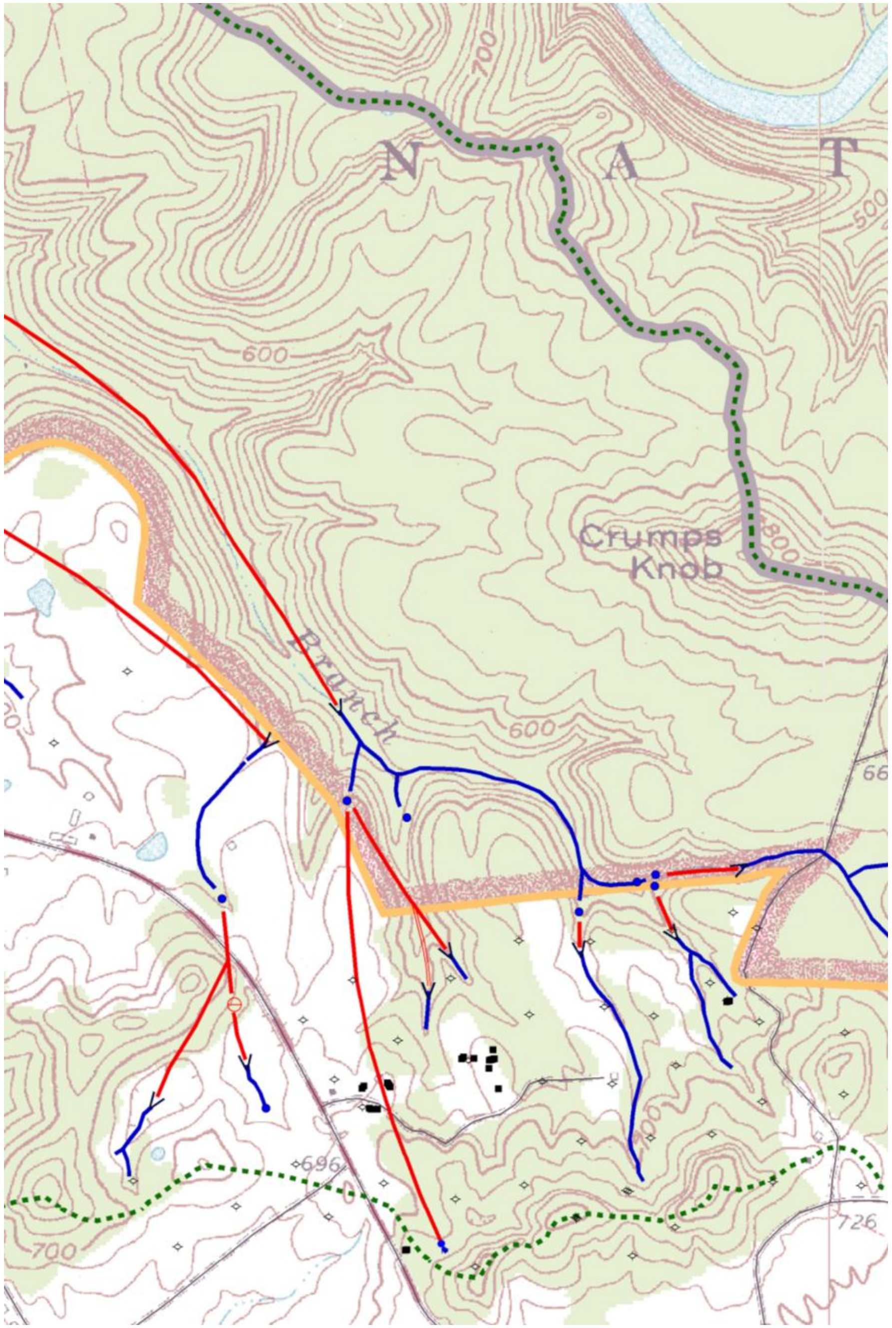
Spatial uncertainty

Scale

PHYSICAL

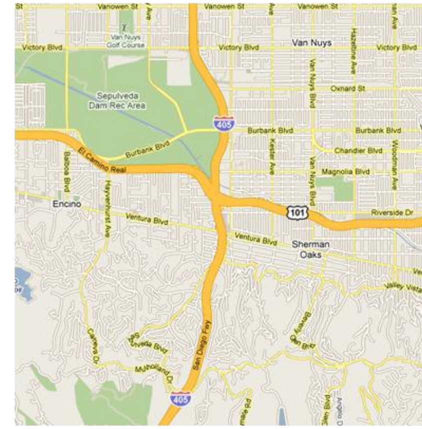
Heterogeneity

Areal interaction



Source: USGS and Glennon

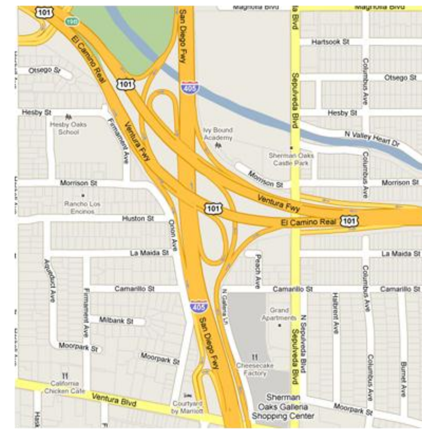
Properties of GENets



REPRESENTATIONAL

Spatial uncertainty

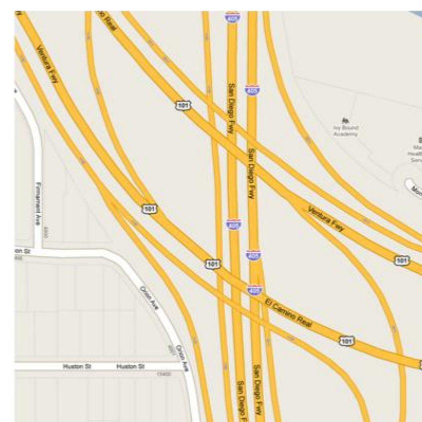
Scale



PHYSICAL

Heterogeneity

Areal interaction



Source: Google Maps 2013

Properties of GENets

Plaza Mayor, Madrid

REPRESENTATIONAL

Spatial uncertainty

Scale

PHYSICAL

Heterogeneity

Areal interaction

Source: Google Maps and Jose Saleta

Properties of GENets

Plaza Mayor, Madrid

REPRESENTATIONAL

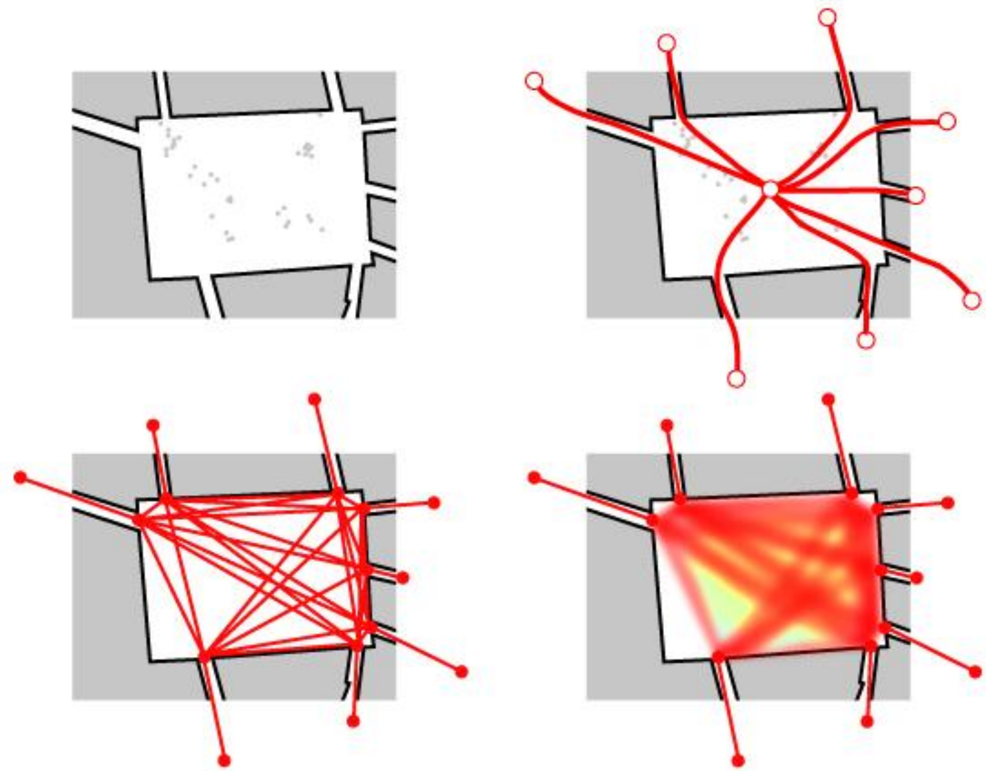
Spatial uncertainty

Scale

PHYSICAL

Heterogeneity

Areal interaction



Properties of GENets

REPRESENTATIONAL

Spatial uncertainty

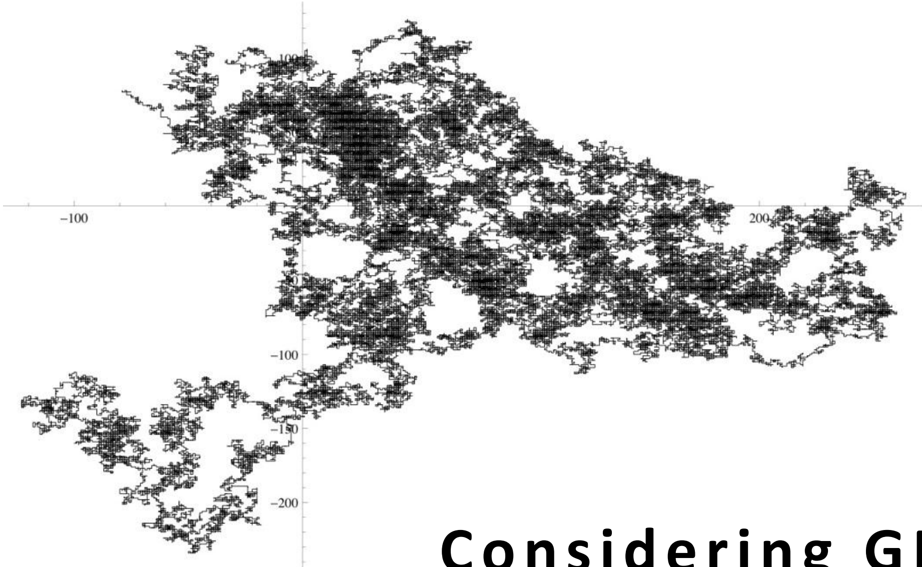
Scale

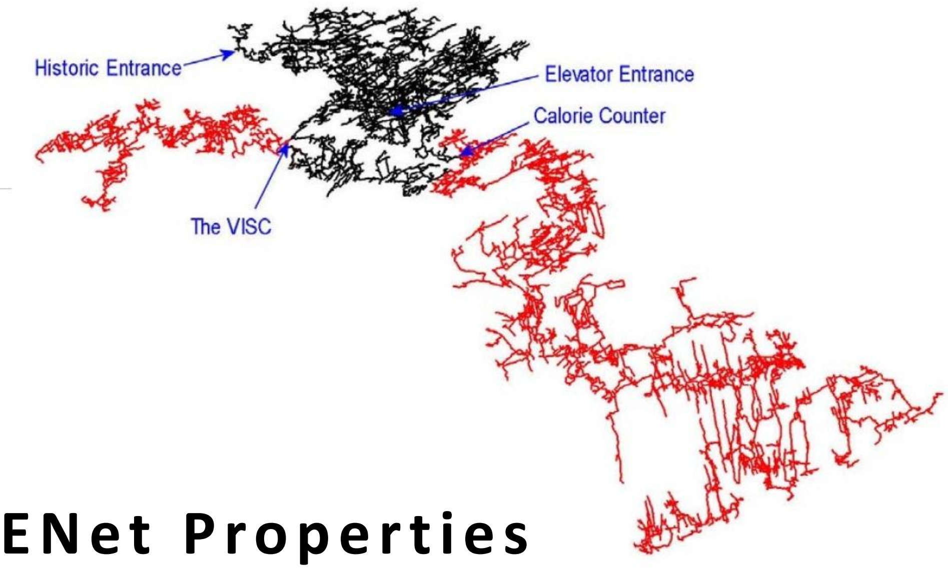
PHYSICAL

Heterogeneity

Areal interaction

http://www.westernwatersheds.org/images/watersheds/eforksalmon-watershed.jpg



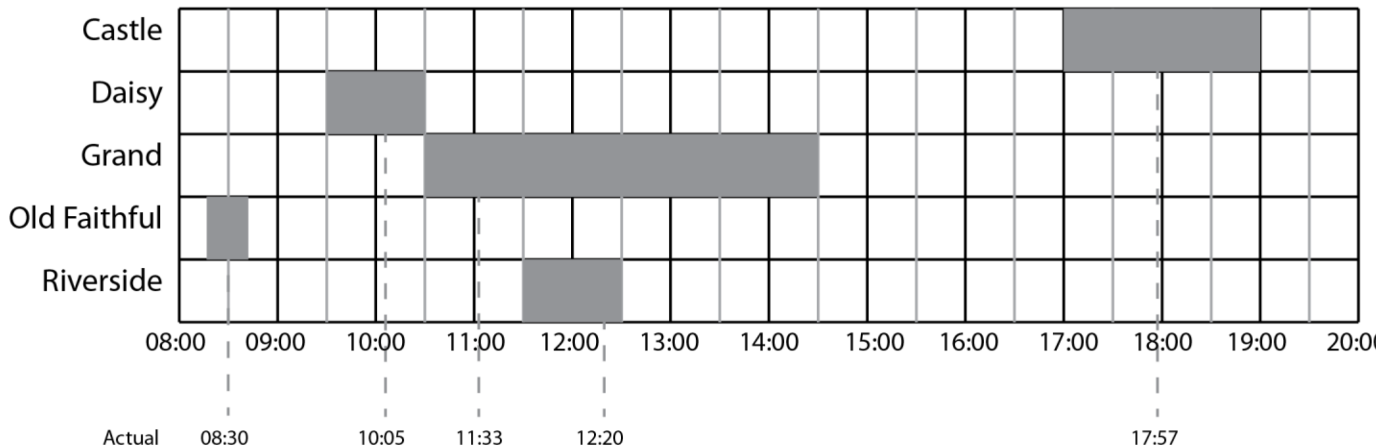


Considering GENet Properties

- Construct more useful and appropriate algorithms: identification of leverageable characteristics, limitations of various analytical approaches
- Development of best practices
- Create maps that afford human understanding of networks

LEFT: A walk seeded by the first 166,000 digits of pi (Davis 2012); RIGHT: map of Jewel Cave, South Dakota (NPS 2008)

July 4, 2012, Prediction Windows




Castle


Daisy


Grand


Old Faithful


Riverside

Itinerary Strategies

Naive

No planning; a random walk through basin

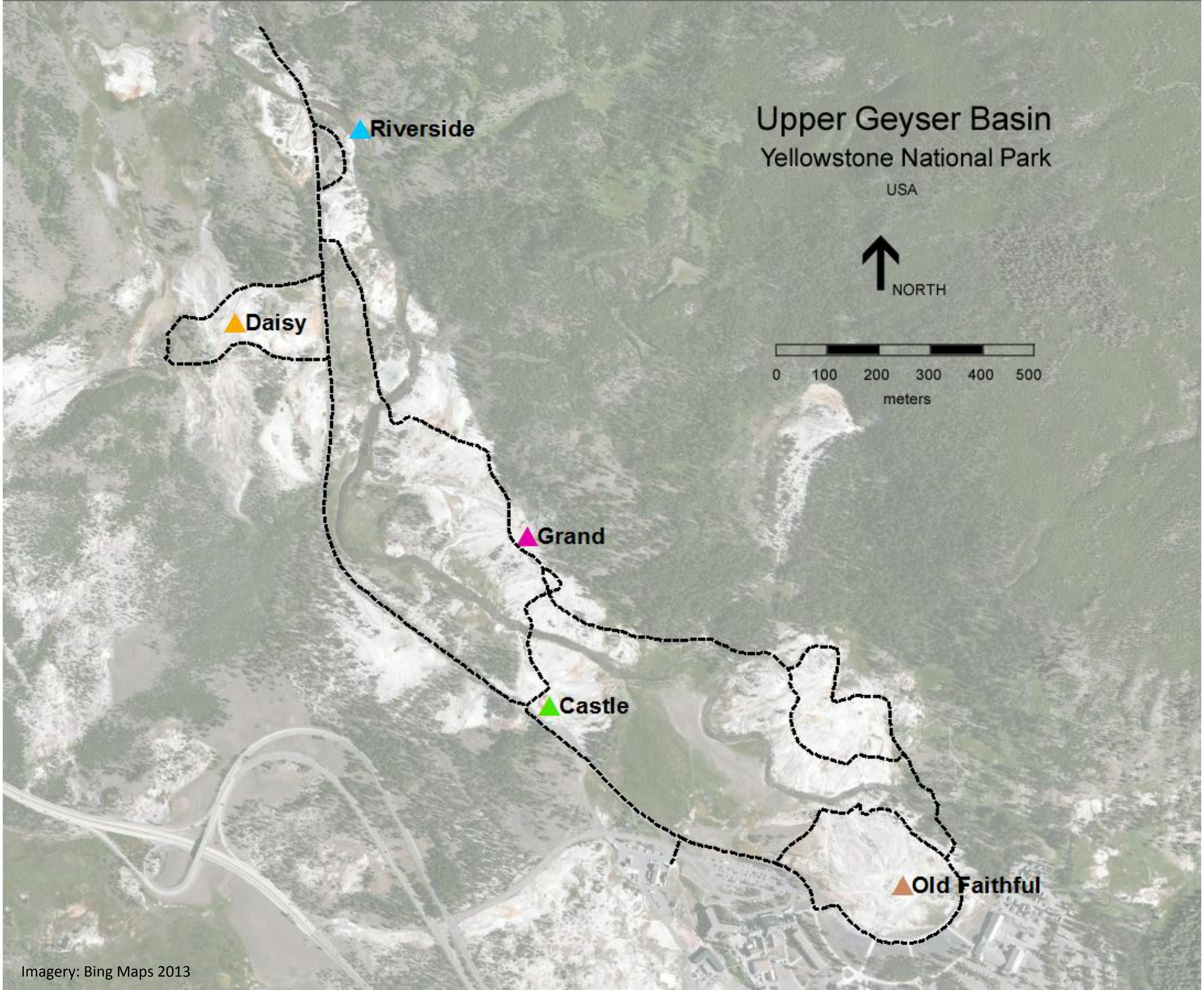
Informed

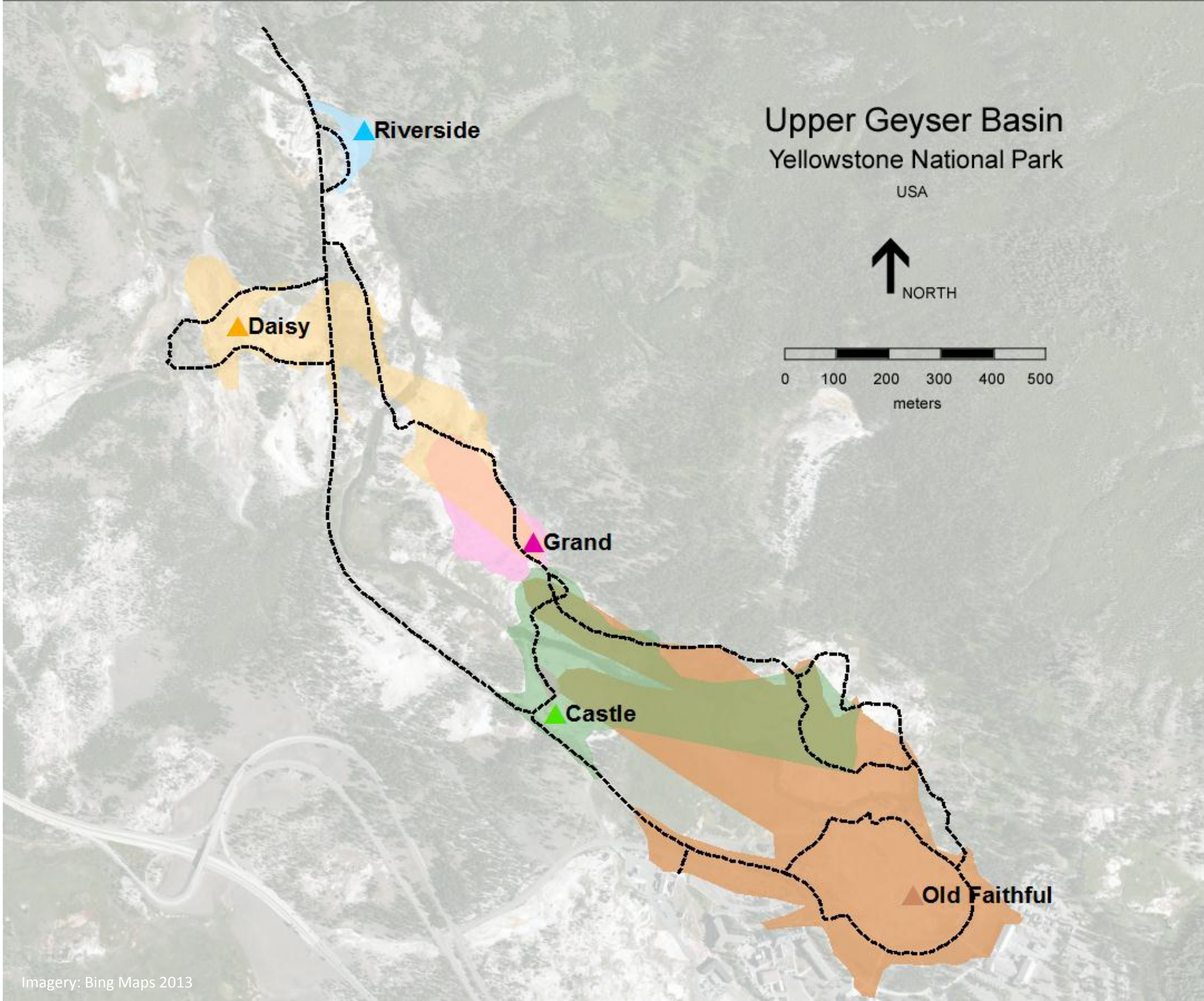
Assess predictions and make an itinerary at the beginning of the day

Expert

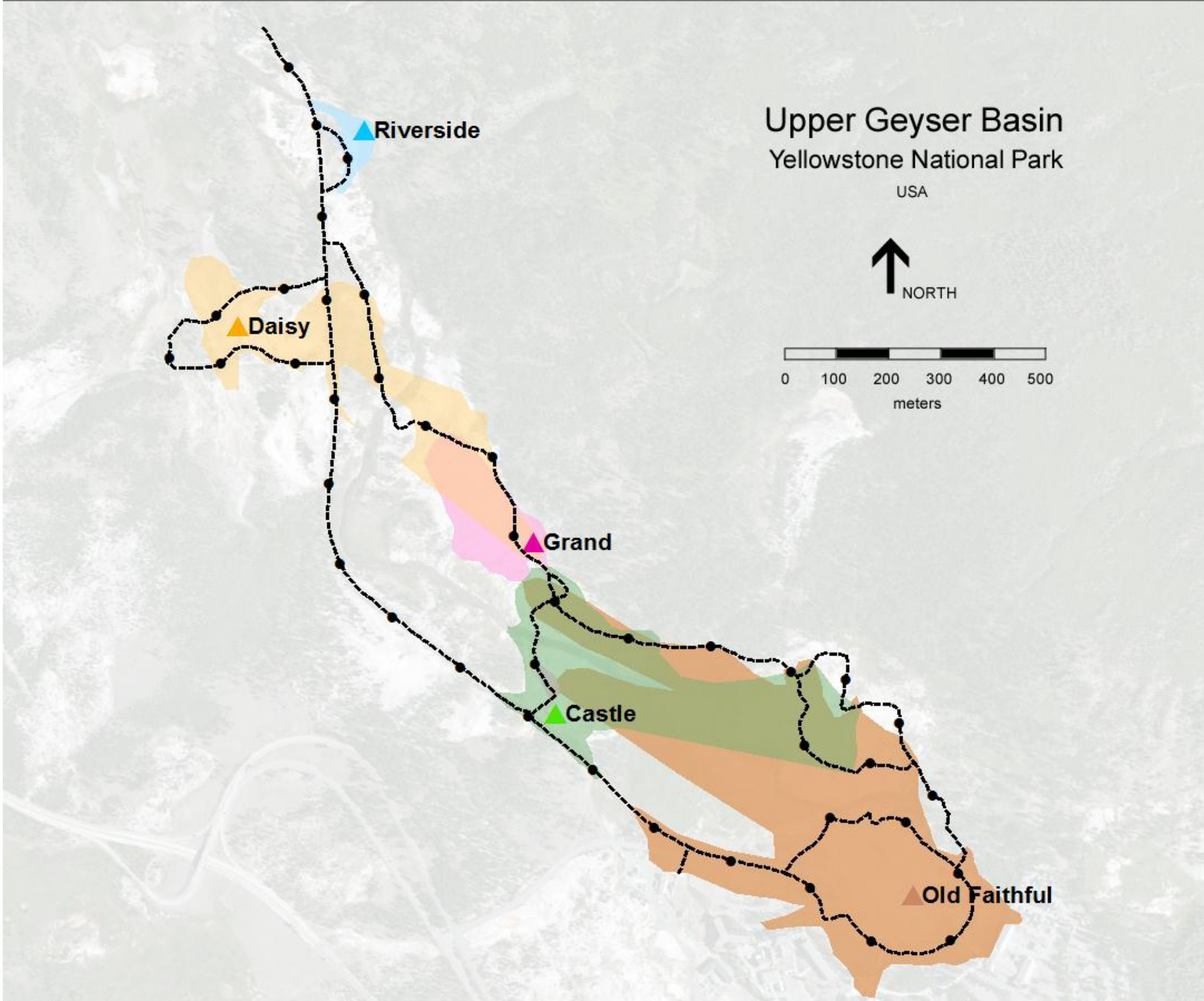
Continuously update itinerary with respect to incoming, real-time data


Upper Geyser Basin
Yellowstone National Park
USA
NORTH
0 100 200 300 400 500
meters
Imagery: Bing Maps 2013

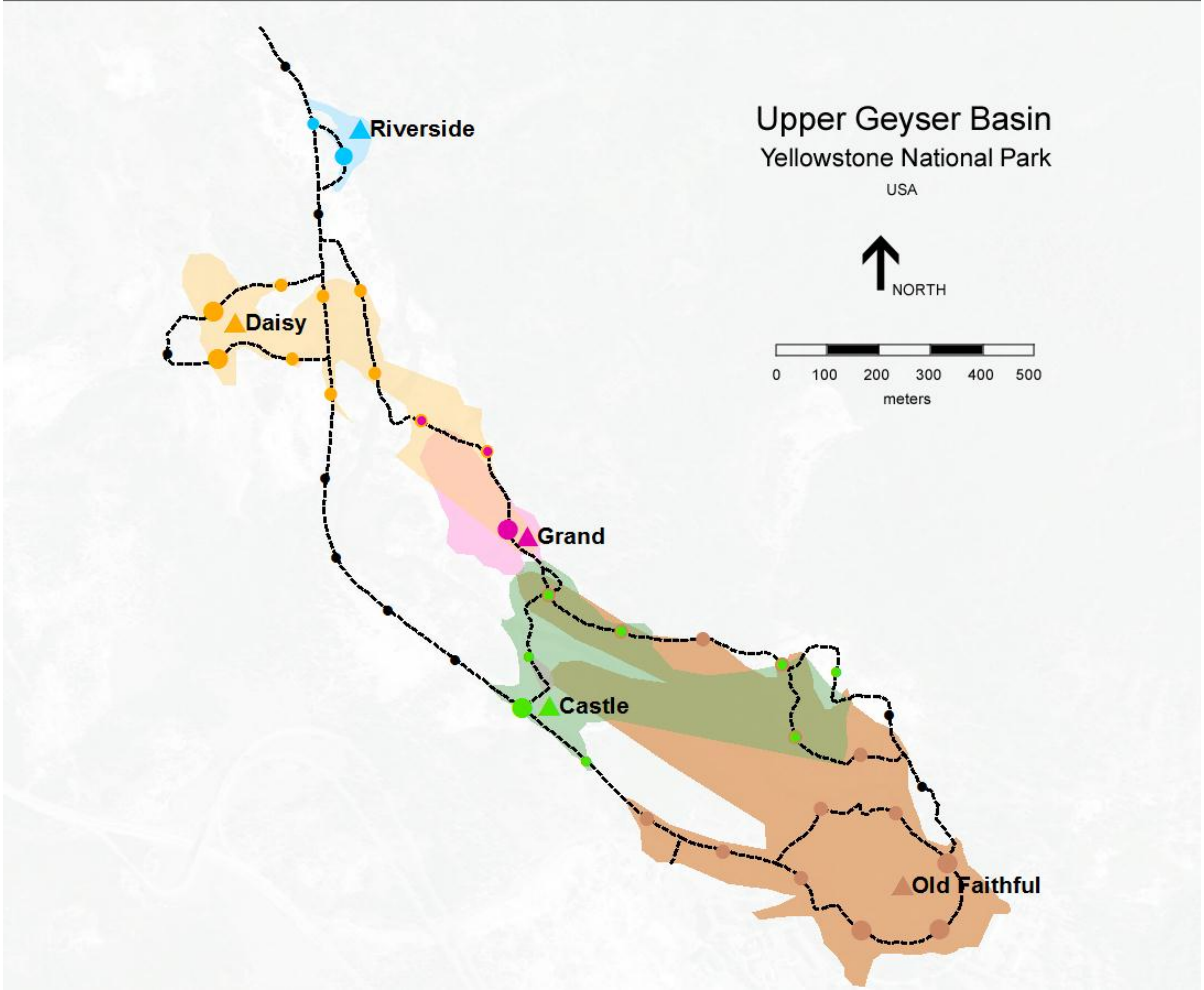

Upper Geyser Basin
Yellowstone National Park
USA
NORTH
0
100
200
300
400
500
meters
Riverside
Daisy
Grand
Castle
Old Faithful
Imagery: Bing Maps 2013



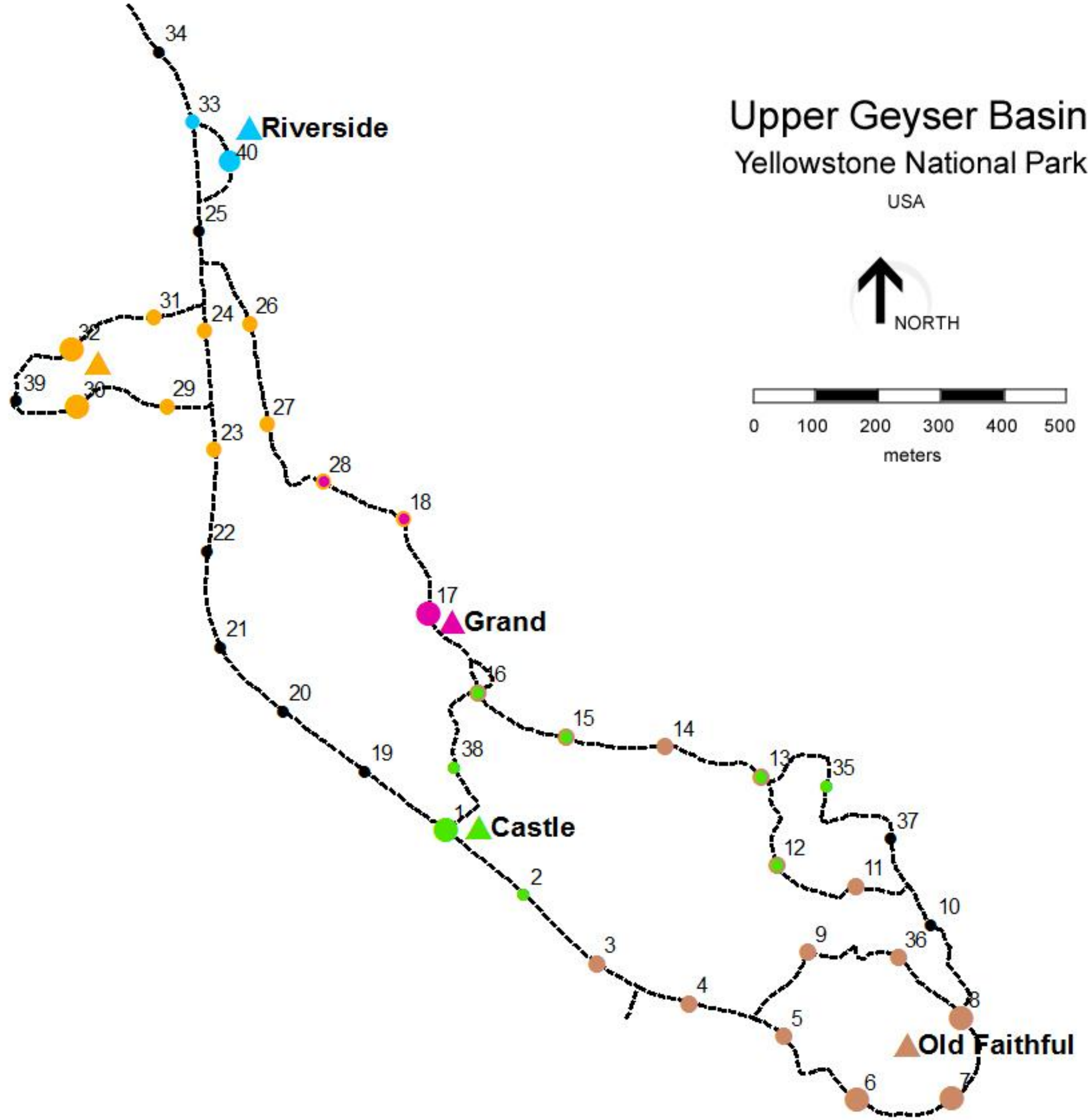
Upper Geyser Basin
Yellowstone National Park
USA
NORTH
0 100 200 300 400 500
meters
Riverside
Daisy
Grand
Castle
Old Faithful
Imagery: Bing Maps 2013



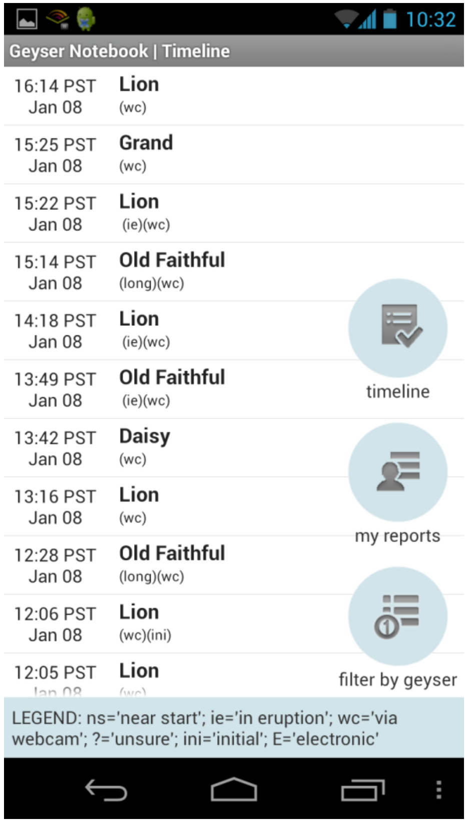
Upper Geyser Basin
Yellowstone National Park
USA
NORTH
0
100
200
300
400
500
meters
Riverside
Daisy
Grand
Castle
Old Faithful

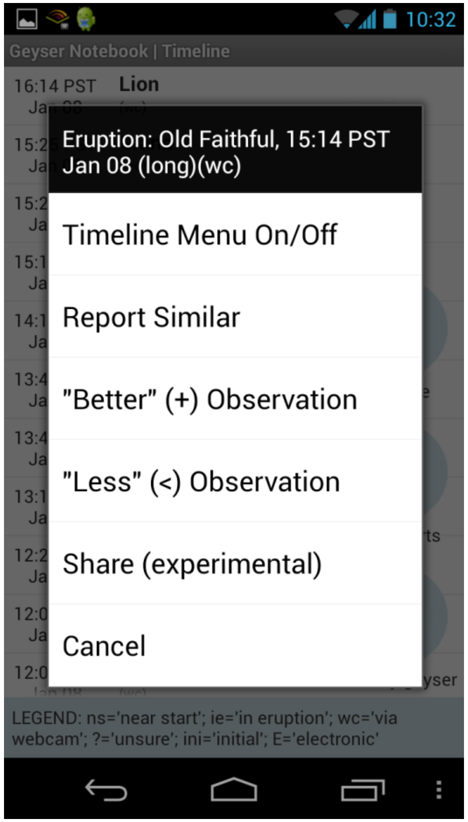


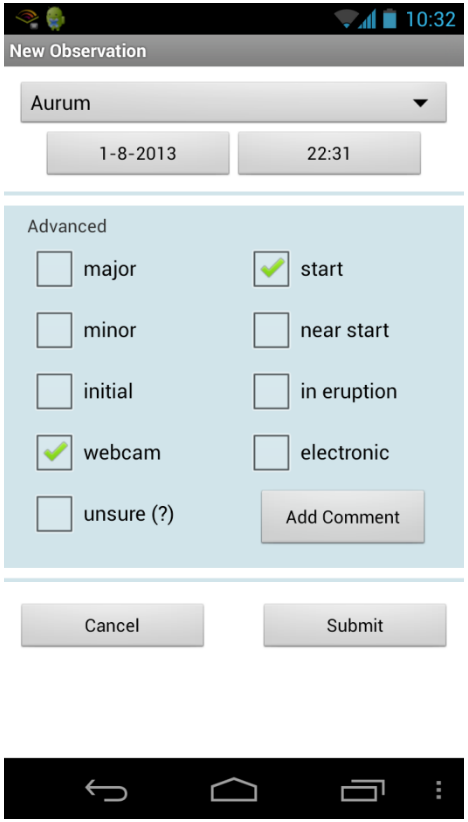
Upper Geyser Basin
Yellowstone National Park
USA
NORTH
0
100
200
300
400
500
meters
Riverside
Daisy
Grand
Castle
Old Faithful

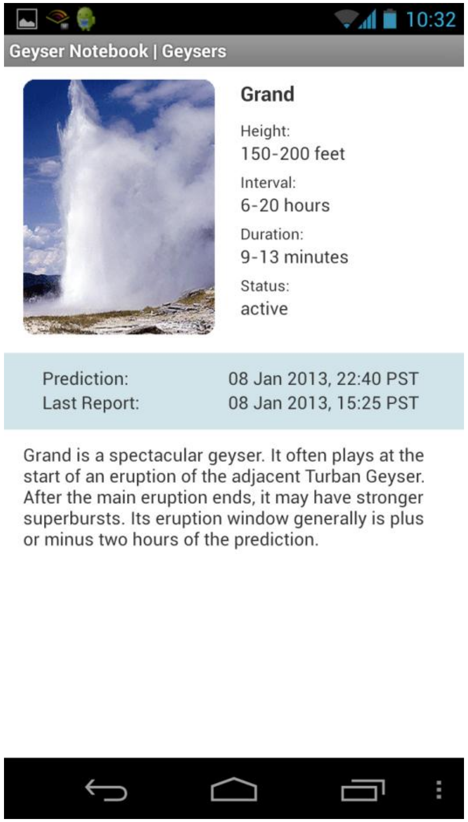

Upper Geyser Basin
Yellowstone National Park
USA
NORTH
0
100
200
300
400
500
meters
Riverside
Grand
Castle
Old Faithful
34
33
40
25
31
24
26
32
39
30
29
27
23
28
18
22
17
21
16
20
15
14
38
19
13
35
37
1
12
11
2
10
9
36
3
4
8
5
6
7

Geyser Notebook

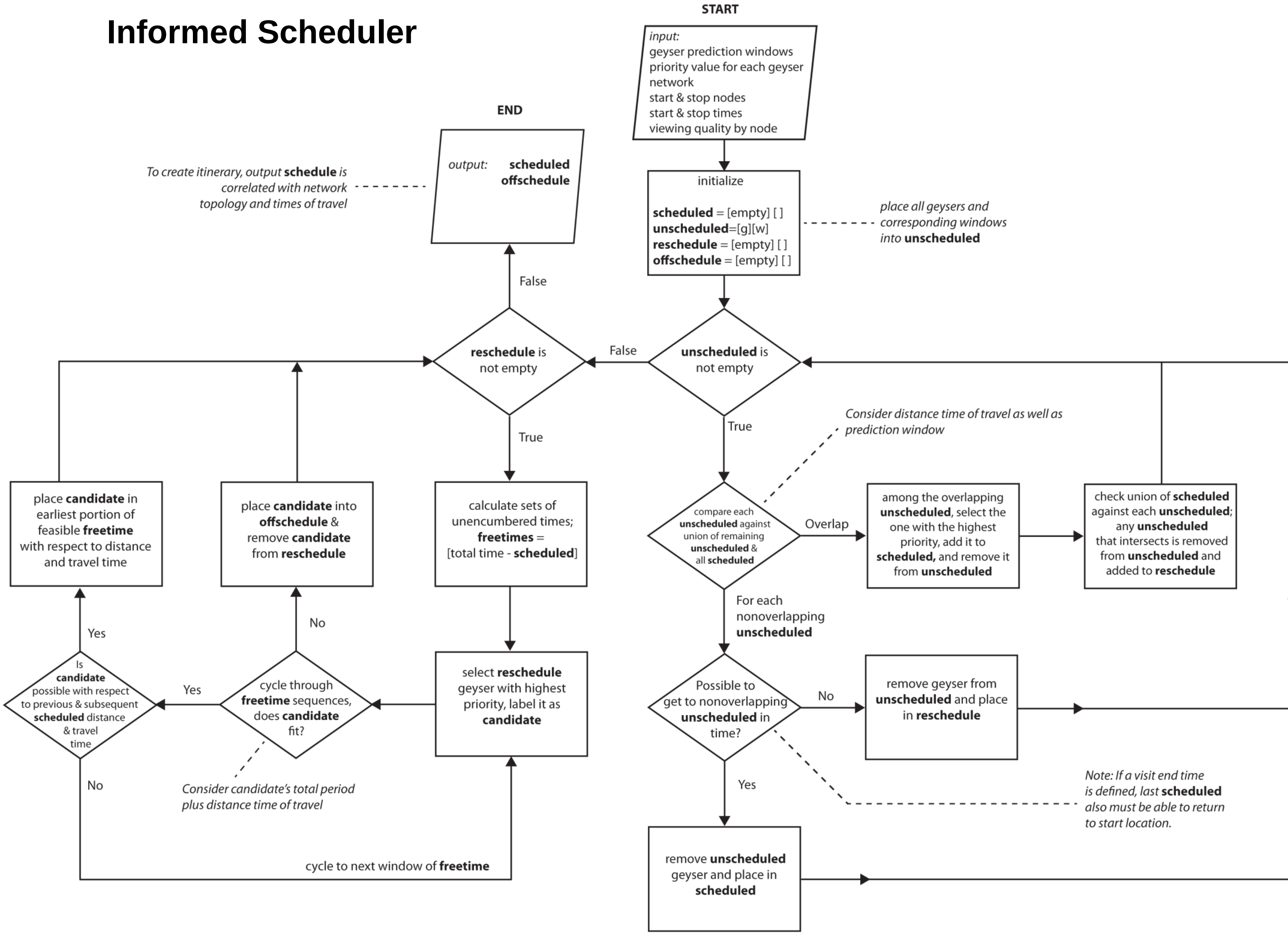




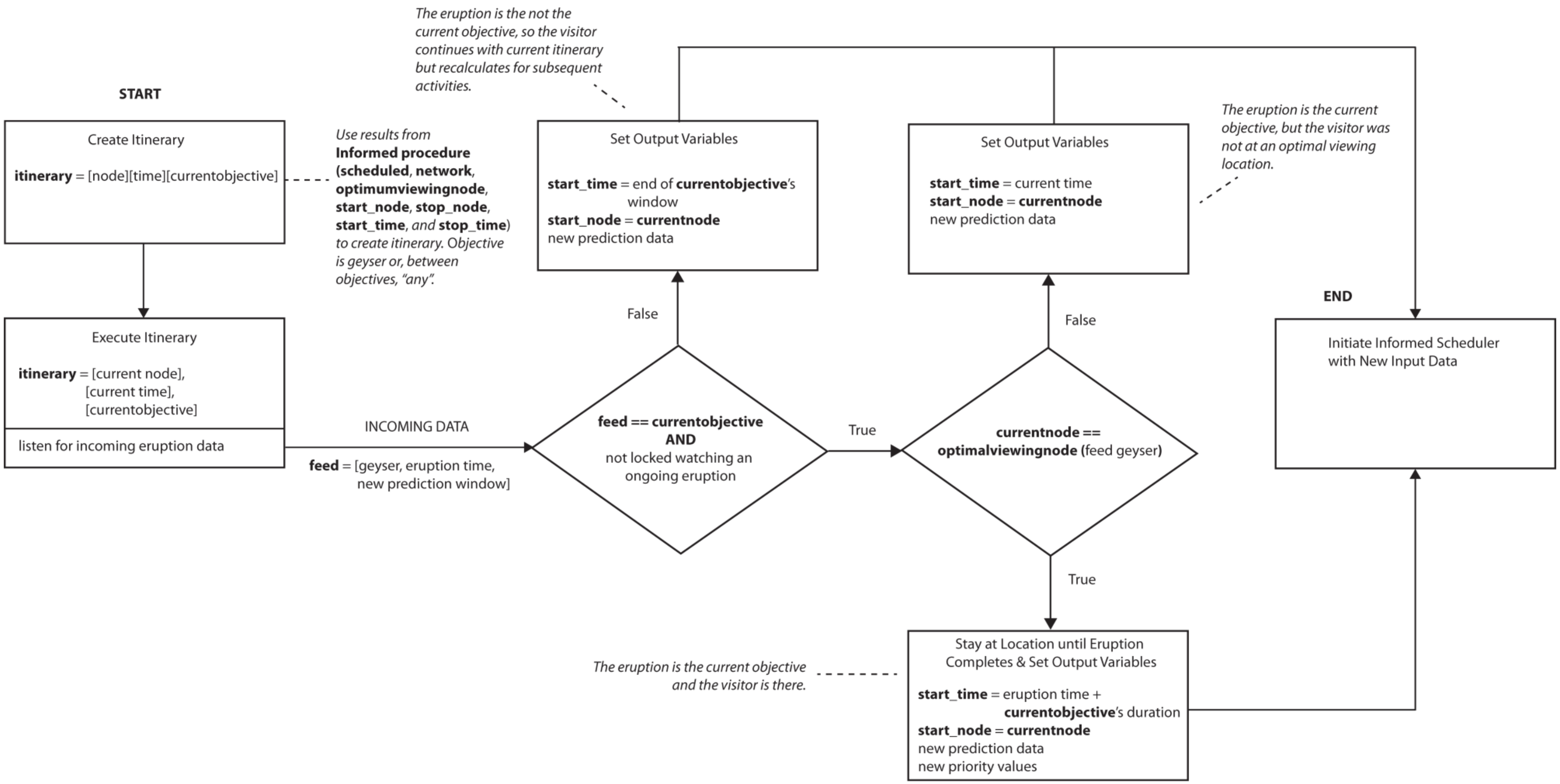




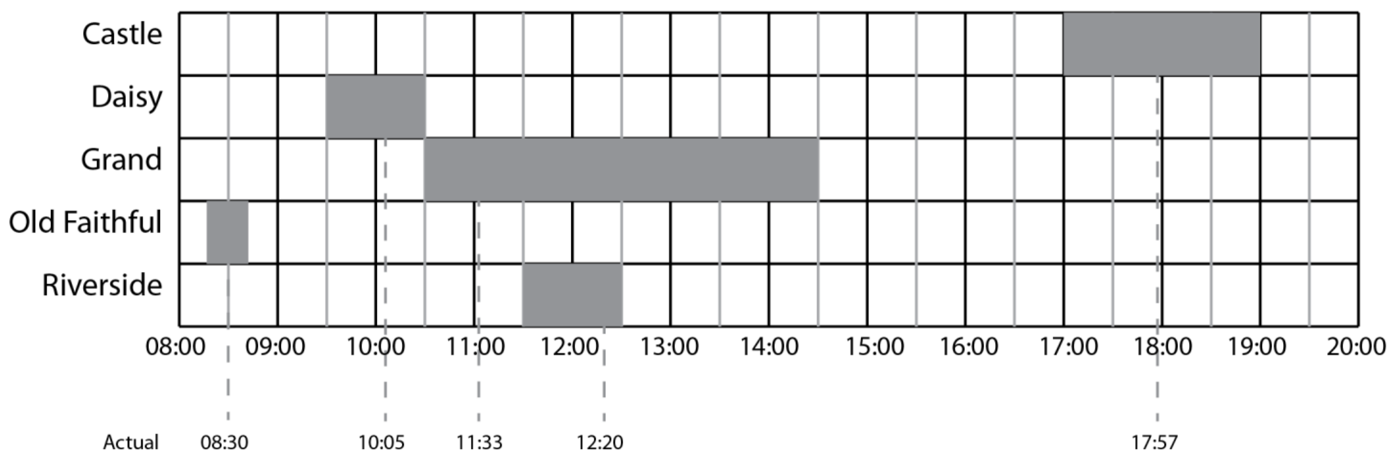
Informed Scheduler



Expert Scheduler



July 4, 2012, Prediction Windows

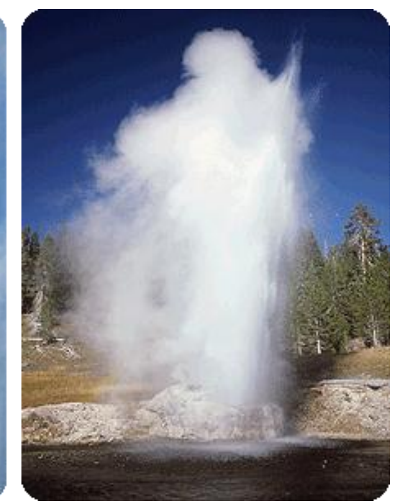


Castle

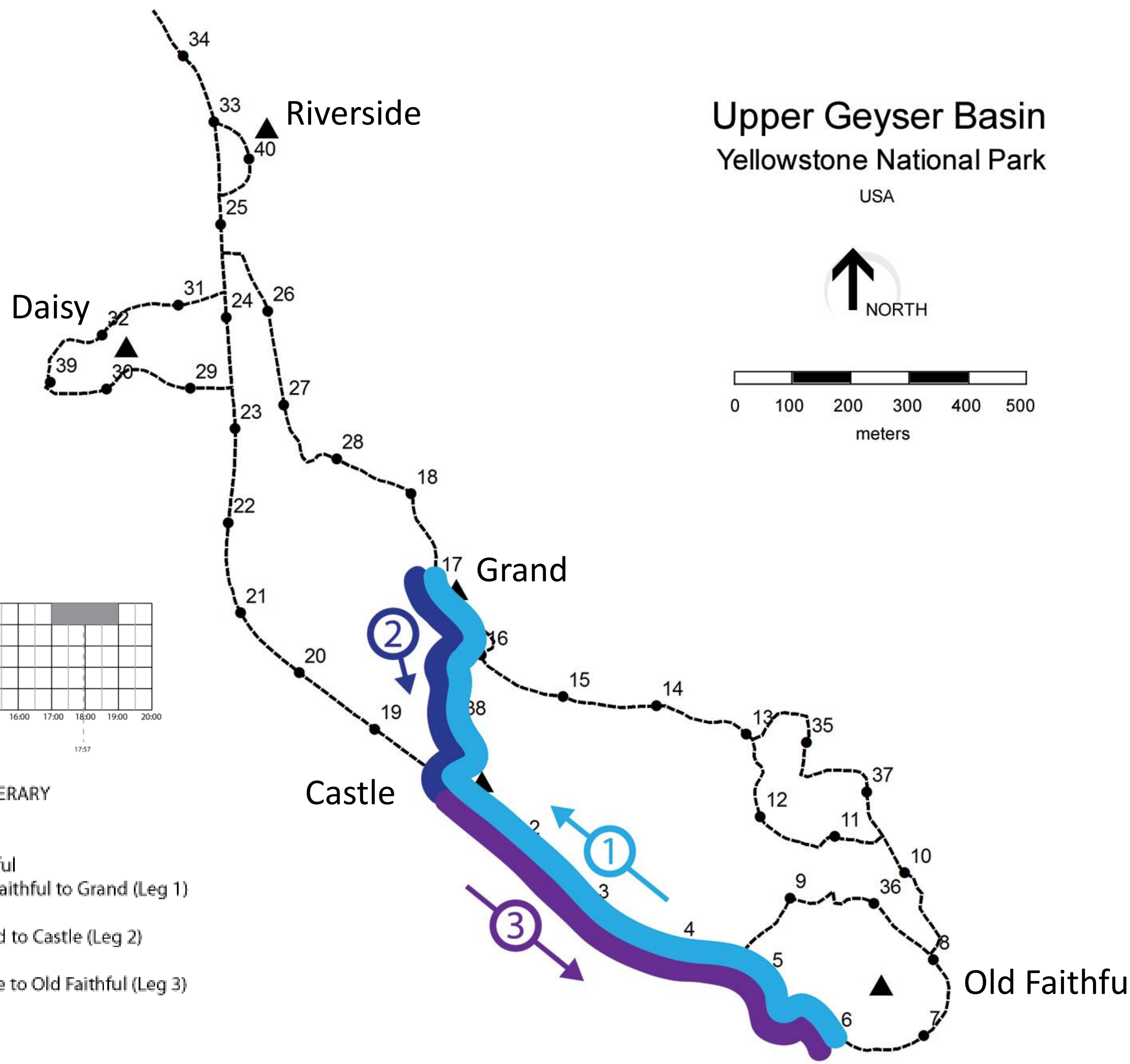
Daisy

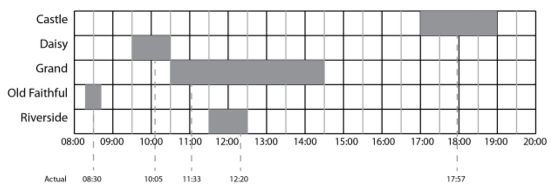
Grand

Old Faithful



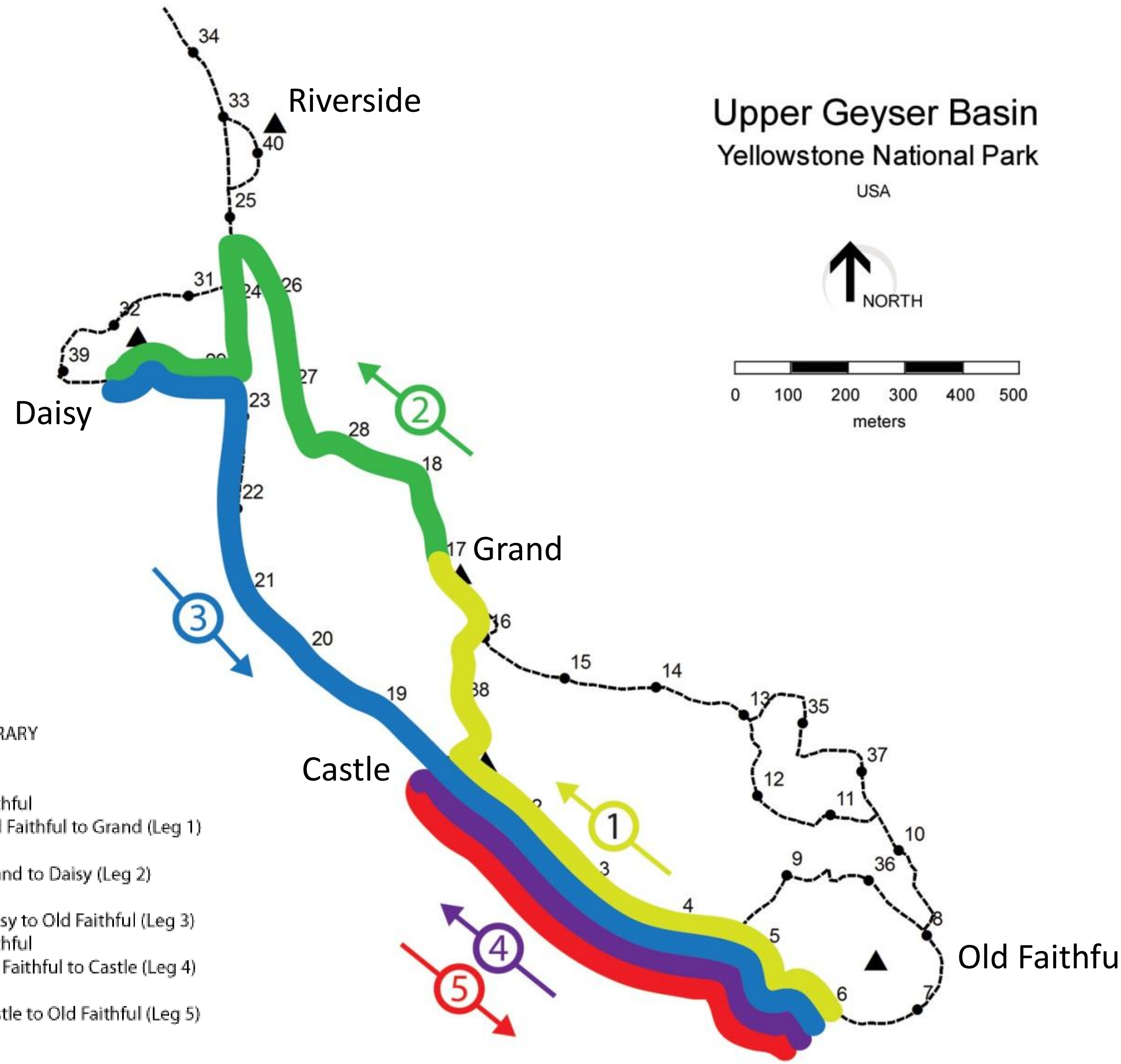
Riverside





INFORMED VISITOR TRAVEL ITINERARY

08:00: Start day at Old Faithful
08:00 to 10:14: Wait at Old Faithful
10:14 to 10:30: Travel from Old Faithful to Grand (Leg 1)
10:30 to 16:54: Wait at Grand
16:54 to 17:00: Travel from Grand to Castle (Leg 2)
17:00 to 19:50: Wait at Castle
19:50 to 20:00: Travel from Castle to Old Faithful (Leg 3)
20:00: End day at Old Faithful



EXPERT VISITOR TRAVEL ITINERARY

08:00: Start day at Old Faithful
08:00 to 10:14: Wait at Old Faithful
10:14 to 10:30: Travel from Old Faithful to Grand (Leg 1)
10:30 to 11:51: Wait at Grand
11:51 to 12:05: Travel from Grand to Daisy (Leg 2)
12:05 to 15:20: Wait at Daisy
15:20 to 15:44: Travel from Daisy to Old Faithful (Leg 3)
15:44 to 16:50: Wait at Old Faithful
16:50 to 17:00 Travel from Old Faithful to Castle (Leg 4)
17:00 to 19:50: Wait at Castle
19:50 to 20:00: Travel from Castle to Old Faithful (Leg 5)
20:00: End day at Old Faithful

GTP Itinerary Results

	Naive with synthetic data	Naive	Informed	Expert
Geyser				
	Mean Observations			
Castle	0 (0.2)	0 (0.2)	1	1
Daisy	1 (1.0)	0 (0.8)	2	2
Grand	0 (0.1)	0 (0.1)	1	1
Old Faithful	4 (4.2)	3 (3.9)	2	3
Riverside	0 (0.1)	0 (0.1)	0	0
	Mean Quality			
Castle	0.6	0.6	1.0	1.0
Daisy	0.6	0.6	0.5	1.0
Grand	0.7	0.7	1.0	1.0
Old Faithful	0.6	0.6	1.0	1.0
Riverside	0.7	0.7	0.0	0.0

Using GENet properties to improve analytical results

SCALE: harmonizing the granularity of the path and events

SPATIAL UNCERTAINTY: data stream verification; “Tragedy of the Data Collection Commons” (missing data); finer scale in important areas to reduce utility ambiguity

AREAL INTERACTION: viewsheds (the neighborhood) affect utility along path; future: leverage spatial autocorrelation of events

CONSTITUENT HETEROGENEITY: possible to achieve more than one objective at the same place; consider other path properties (like path width and wind direction)

Contributions to the Field

- The identification of an initial set of GENet characteristics and description of methods for uncovering more such properties

- A technique for creating geographic data models

- The development of a geographic data model for GENet flow; formalization of known and uncertain linkages with respect to flow

- The development of a GIS workflow for GENet itinerary creation that addresses: recurrent visits and satiation; multiple objectives at a single location; and real-time activities with uncertain completion times

What's Next?

- Look for more GENet properties; find people to help
- Explore *methods* for integrating dynamic GENet attributes
- Characterize Abstract GENets
- Describe linear programming formulation for optimal solution to GTP
- Generalize GTP to other itinerary creation domains
- Explore the nature of geyser eruption moving means
- Continue exploration of real time mobile decision making (egocentric spatial analysis)
- Enumerate and describe a "Naïve Geography" of networks in geography
- Study, describe, and explore alternatives to overcome the "Tragedy of the data collection commons"
- Participate in time comparison research in GIS
- Create, use, and encourage general optimization tools for GIS

Acknowledgements

Mike Goodchild
Rick Church
Keith Clarke
Shih-Lung Shaw
Rhonda Glennon
Josh Bader
Karl Grossner
Indy Hurt
Helen Couclelis
Sara Fabrikant
James Frew
Don Janelle
Hugo Loaiciga
Dan Montello
Waldo Tobler
Matt Rice
John Gallo
Andrea Nuernburger
Sean Benison
Jorge Sifuentes
Linna Li
Matt Vitale

Kitty Currier
Reg Archer
Carlos Baez
Tim Niblett
Matt Niblett
Julie Dillemuth
Laurel Sutter
Wenn Wen Li
Shaun Walbridge
Chad Catacchio
Kailen Wright
Eric Davila
Suzanne Foss
Keely Roth
Nate Royal
Pam Dalal
Jeff Howarth
Ed Pultar
Kerry Halligan
Susan Tran
Tom Pingel
Drew Dara-Abrams

Bryan Karaffa
Guylene Gadal
Dylan Parenti
Karen Doehner
Mo Lovegreen
Connie Padilla
Susanna Baumgart
Beilei Zhang
Jose Saleta
Bernadette Weinberg
Will McClintock
Tom Cova
May Yuan
Dan Sui
Diana Sinton
Karen Kemp
Ben Zhao
Barbara Harthorn
Fiona Goodchild
Jack Dangermond
Laura Dangermond

Jacob Young
Graham Meech
Tara Cross
Jeff Cross
Lynn Stephens
Mary Beth Schwartz
Scott Bryan
Clark Murray
Vicki Whitledge
David Goldberg
Dean Lohrenz
Jim Scheirer
Kitt Barger
Ralph Taylor
Ashley Glennon
Suzanne Glennon
Merritt Glennon
Christian Glennon
Keith Pfaff
Arlene Pfaff
Bob Glennon
Rita Glennon