

Predicting Map Error by modeling the Sacramento River Floodplain

Joshua H. Viers^{1,2}

Alexander K. Fremier³

and Rachel Anaïs Hutchinson¹



1 Information Center for the Environment, Department Environmental Science & Policy,
University of California, Davis

2 Center for Watershed Sciences, University of California, Davis

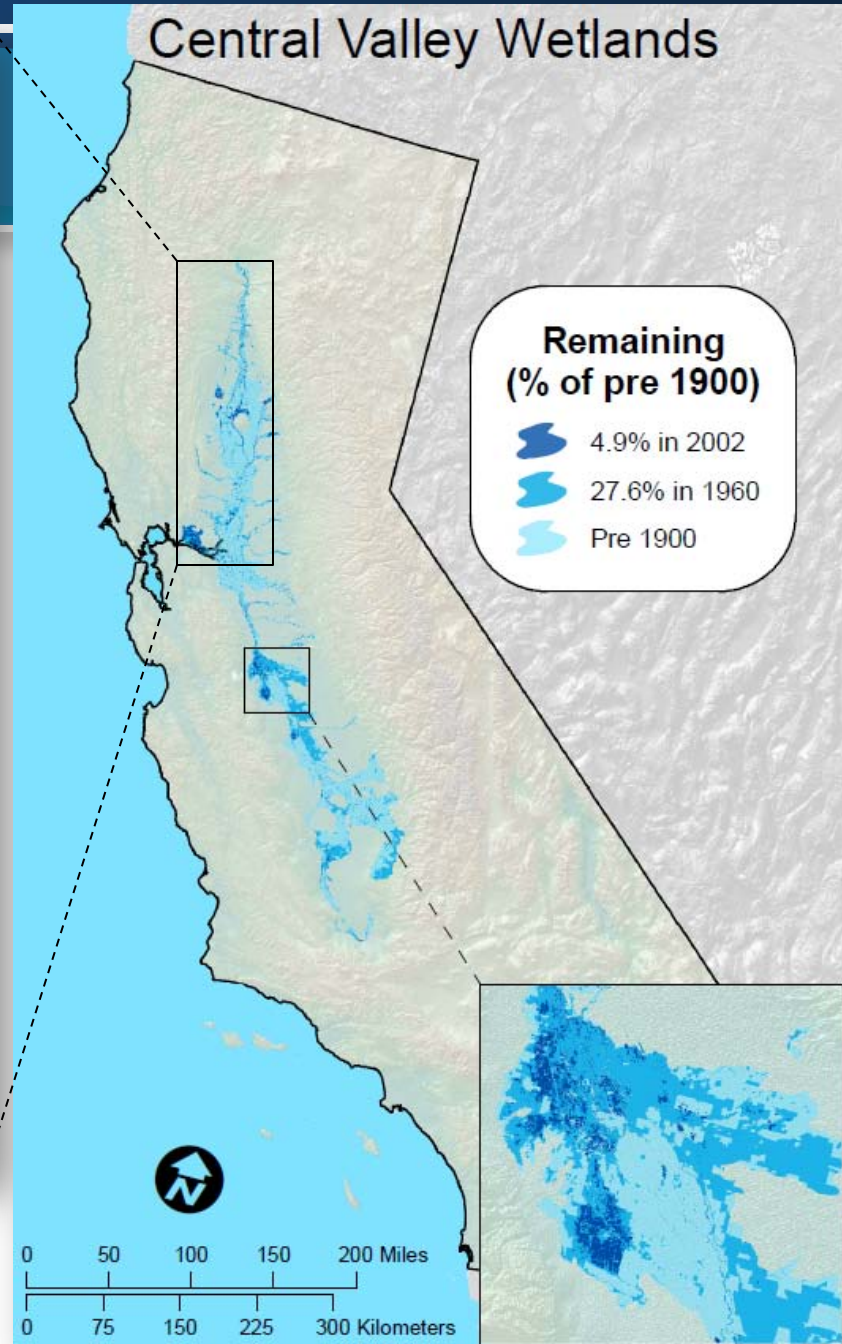
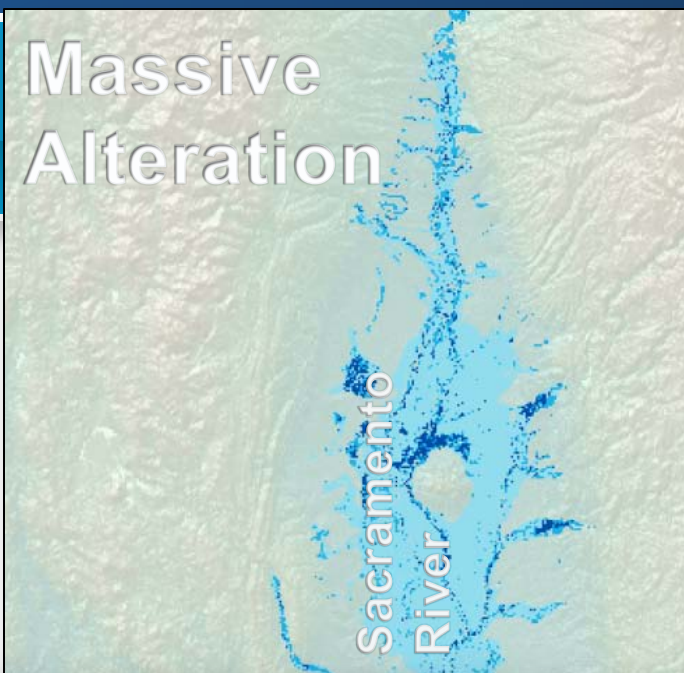
3 College of Natural Resources, University of Idaho

jhviers@ucdavis.edu

2010 ESRI International User Conference, San Diego, California



Massive Alteration





Sacramento River Vegetation Map

CSU Chico GIC (Nelson et al.) 2007

~ 100 miles of RIVER

~ 60,000 acres of VEGETATION

~ 8000 polygons in MAP



Pine Creek Example





Pine Creek Example





Accuracy Assessment



3-tiered Approach

- Geolocated Rapid Assessments
- Visual Check Using ArcPad
- Independent Digitization





Accuracy Assessment



Rapid Assessments

The rapid assessment protocol (RA) was developed by:

California Native Plant Society (cnps.org)

and California Fish & Game (dfg.ca.gov)

as a standardized method to quickly assess and map vegetation types over relatively large, ecologically defined regions. Rapid assessments are used to determine ecological variation across landscapes, habitat composition, and site quality.



Accuracy Assessment



Rapid Assessments

We collected rapid assessments in areas that were not well represented by existing map units or defined vegetation classes.

Examples:

Herbaceous types: mugwort (*Artemisia douglasiana*)

Forest types: white alder (*Alnus rhombifolia*)

Invasive species: fig (*Ficus carica*)



Accuracy Assessment



Visual Check using ArcPad

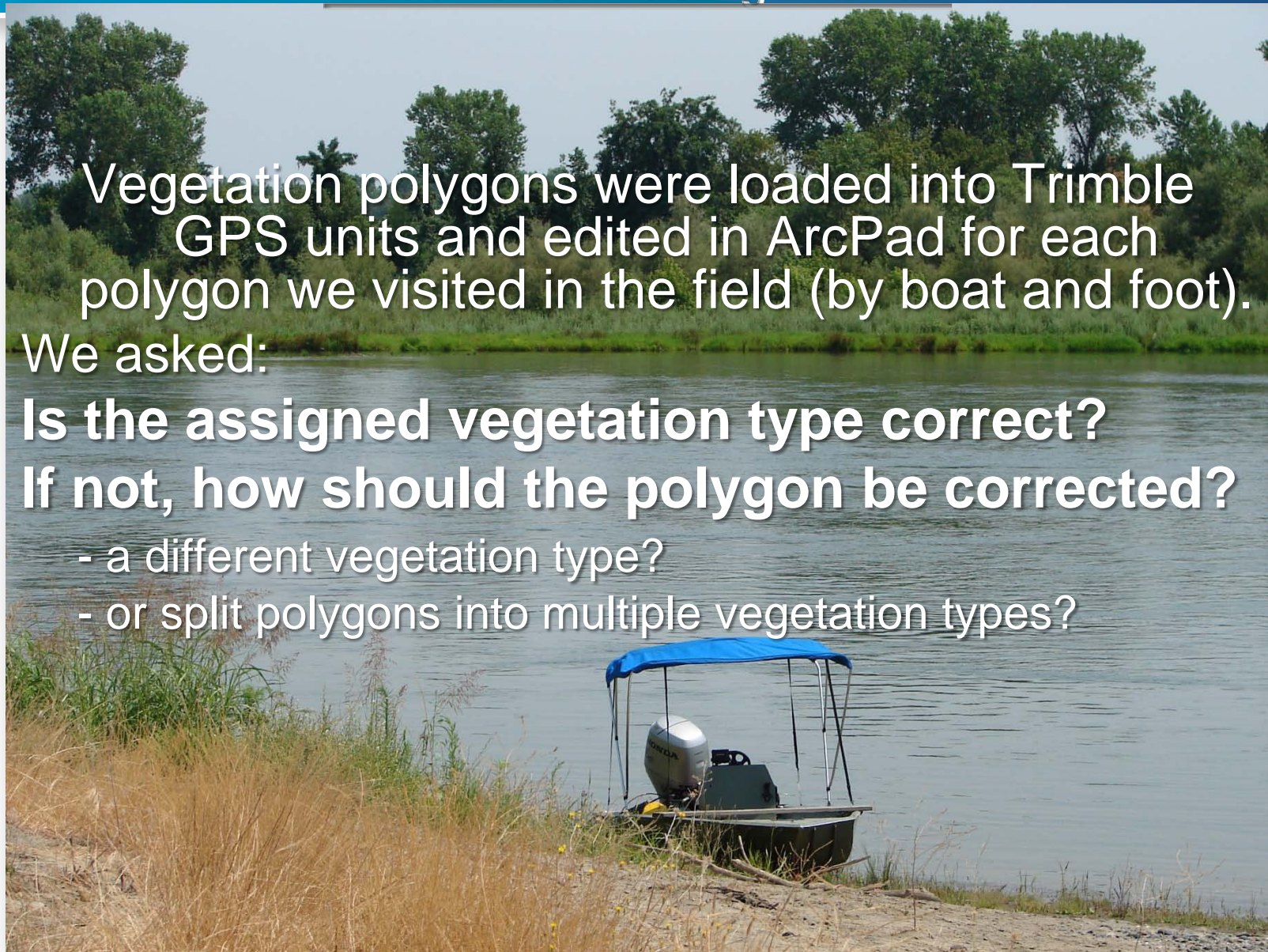
Vegetation polygons were loaded into Trimble GPS units and edited in ArcPad for each polygon we visited in the field (by boat and foot).

We asked:

Is the assigned vegetation type correct?

If not, how should the polygon be corrected?

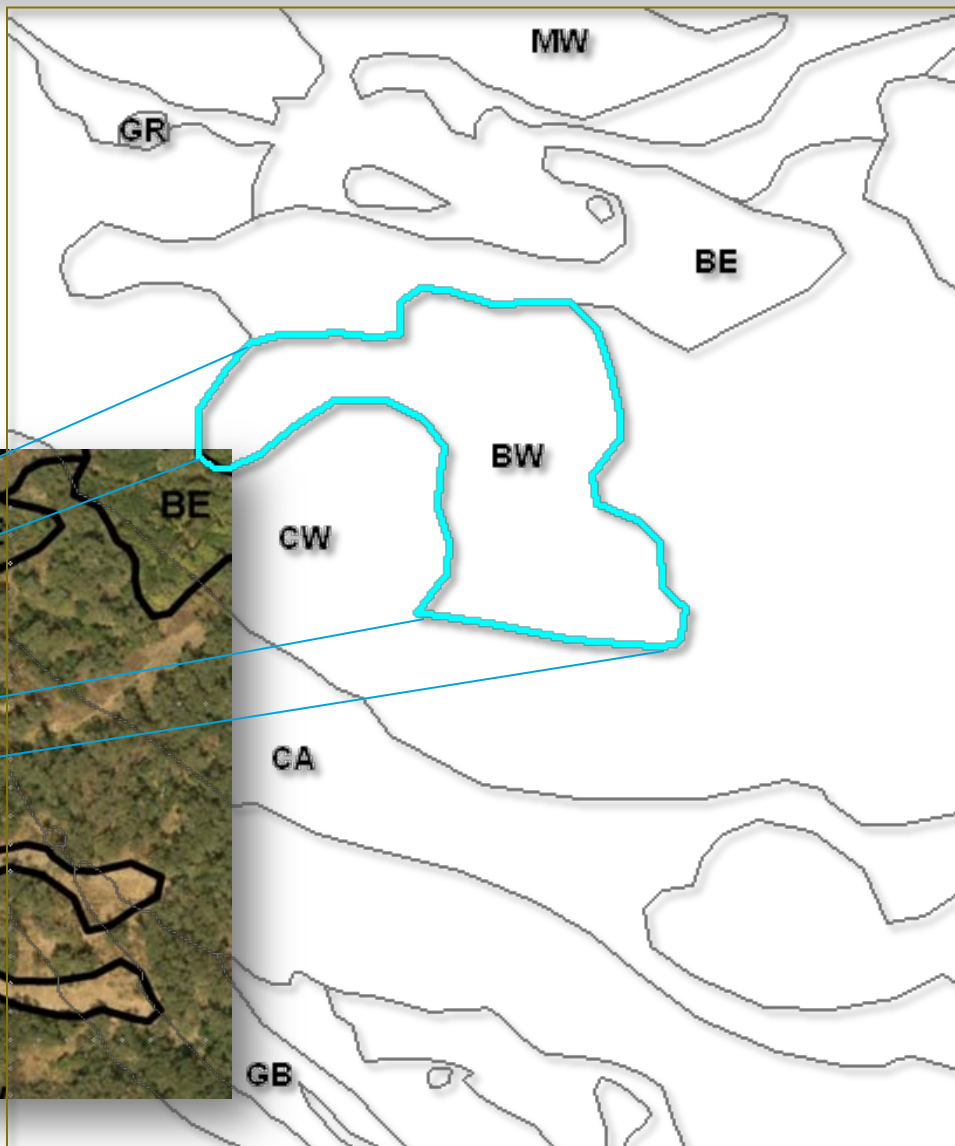
- a different vegetation type?
- or split polygons into multiple vegetation types?





Accuracy Assessment

Visual Check using ArcPad



Map Legend

BE: box elder

BW: black walnut

CA: annual
grassland

CW: cottonwood

GB: gravel bar

GR: giant reed

MW: mixed willow



Accuracy Assessment



Results

Field Effort:

~ 15% of the total number of polygons (1,227)

>10% of all forest or herbaceous polygons

<10% of BS (blackberry),
GR (giant reed),
LP (*Ludwigia*), and
OW (open water)

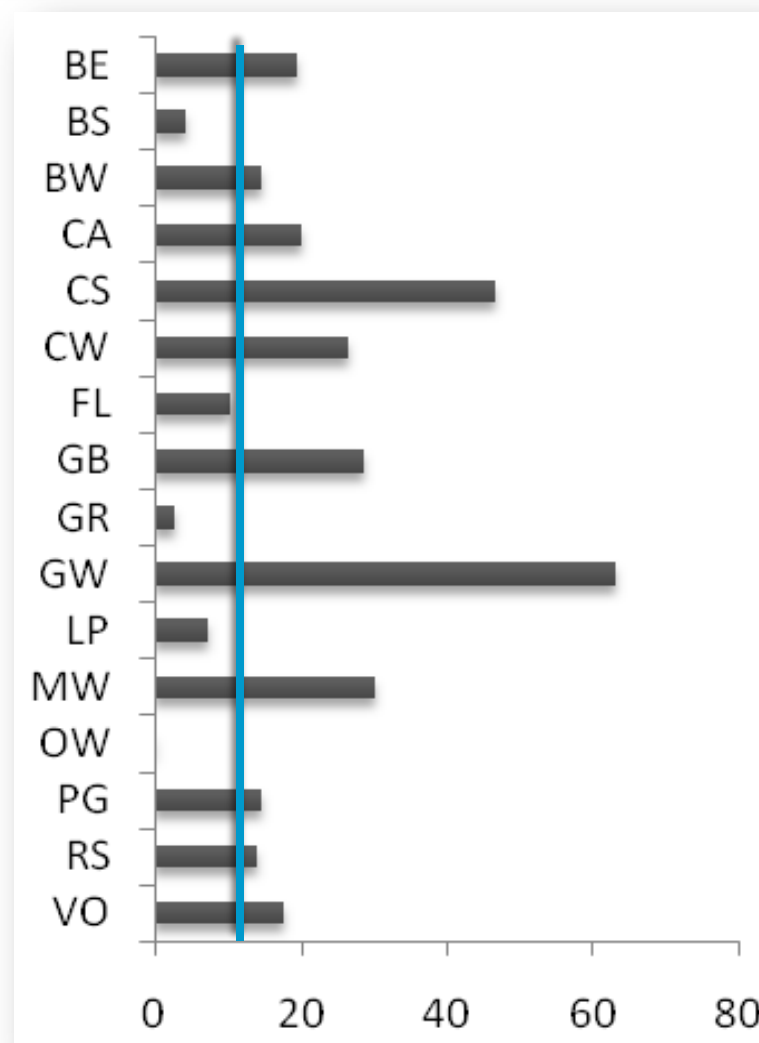




VegMap Field Check



Code	Category
BE	Box Elder
BS	Blackberry Scrub
BW	CA Black Walnut
CA	CA Annuals
CS	CA Sycamore
CW	Fremont Cottonwood
FL	Floating Leaf
GB	Gravel Bar
GR	Giant Reed (<i>Arundo donax</i>)
GW	Goodding's Willow
LP	<i>Ludwigia peploides</i>
MW	Mixed Willow
OW	Open Water
PG	Introduced Perennials
RS	Riparian Scrub
VO	Valley Oak



Percent of Polygons Field Checked



Accuracy Assessment

Rapid Assessments + Visual Check using ArcPad Results



BE: box elder
 BS: blackberry scrub
 BW: black walnut
 CA: annual grassland
 CS: California sycamore
 CW: cottonwood

Count	BE	BS	BW	CA	CS	CW
Col %						
Row %						
BE	63	0	3	0	0	2
	82.89	0	3.41	0	0	0.86
	80.77	0	3.85	0	0	2.56
BS	0	8	0	0	0	0
	0	80	0	0	0	0
	0	88.89	0	0	0	0
BW	6	0	73	0	0	1
	7.89	0	82.95	0	0	0.43
	6.98	0	84.88	0	0	1.16
CA	0	0	0	95	0	0
	0	0	0	81.2	0	0
	0	0	0	95	0	0
CS	1	0	2	0	22	2
	1.32	0	2.27	0	100	0.86
	2.56	0	5.13	0	56.41	5.13
CW	2	0	3	0	0	214
	2.63	0	3.41	0	0	91.85
	0.84	0	1.26	0	0	89.54



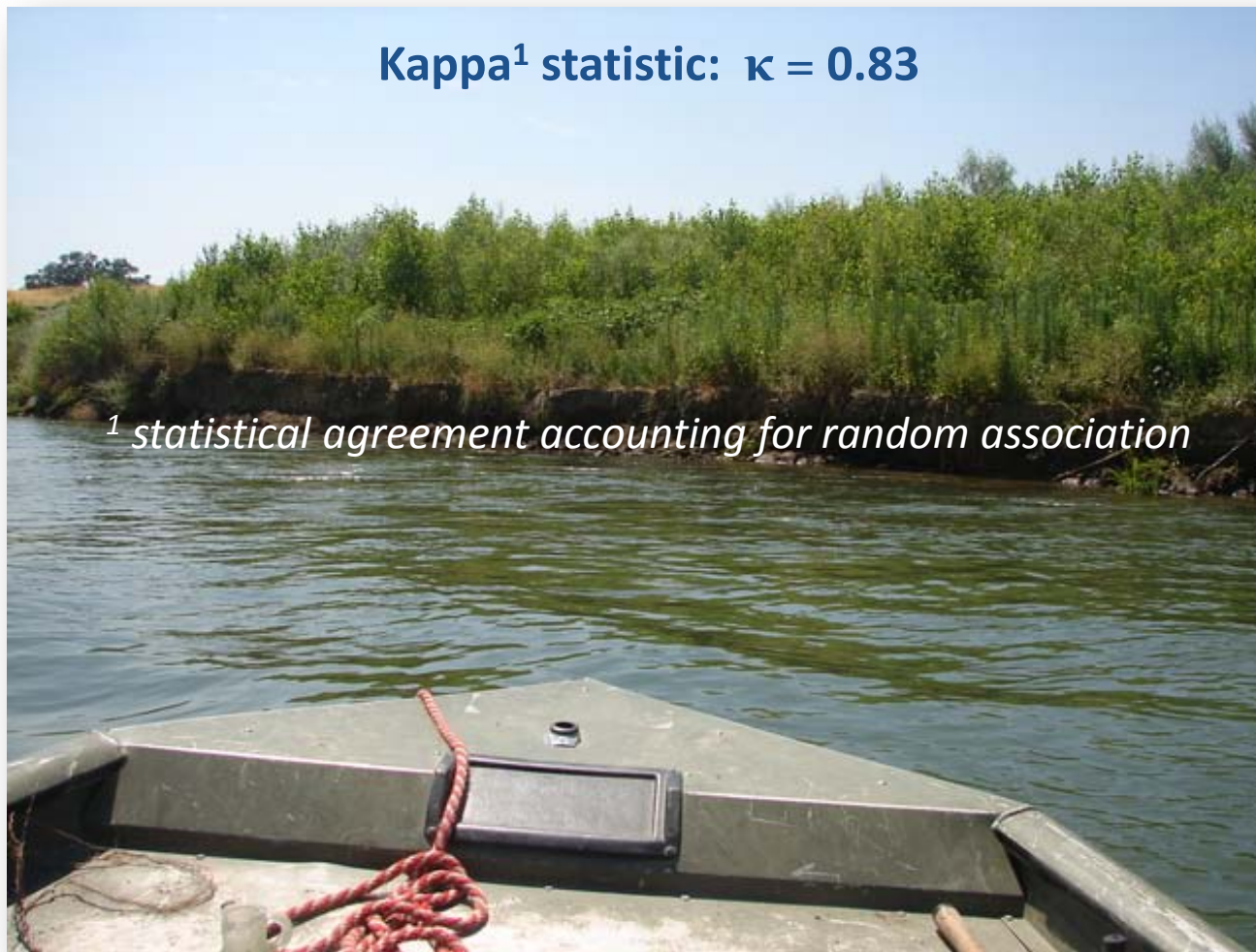
Accuracy Assessment

Rapid Assessments using ArcPad
Results



Total Accuracy: 85.3%

Kappa¹ statistic: $\kappa = 0.83$



¹ statistical agreement accounting for random association



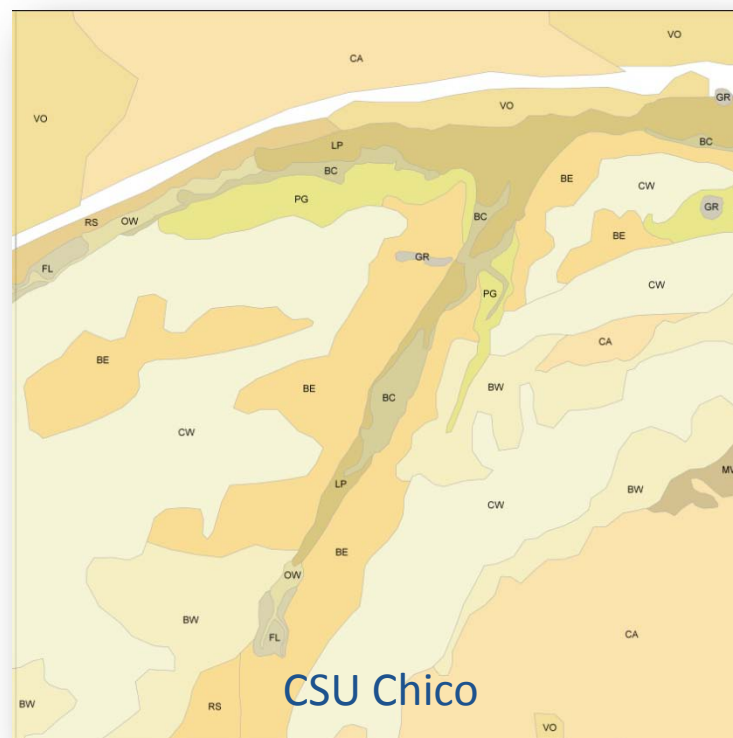
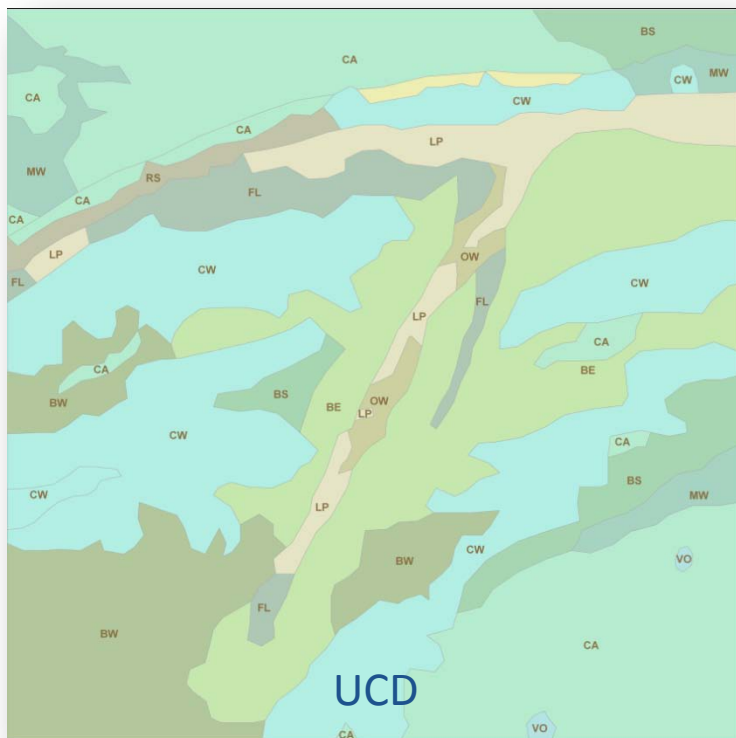
Accuracy Assessment

Independent Digitization



Re-digitized 500m by 500m blocks of riparian vegetation

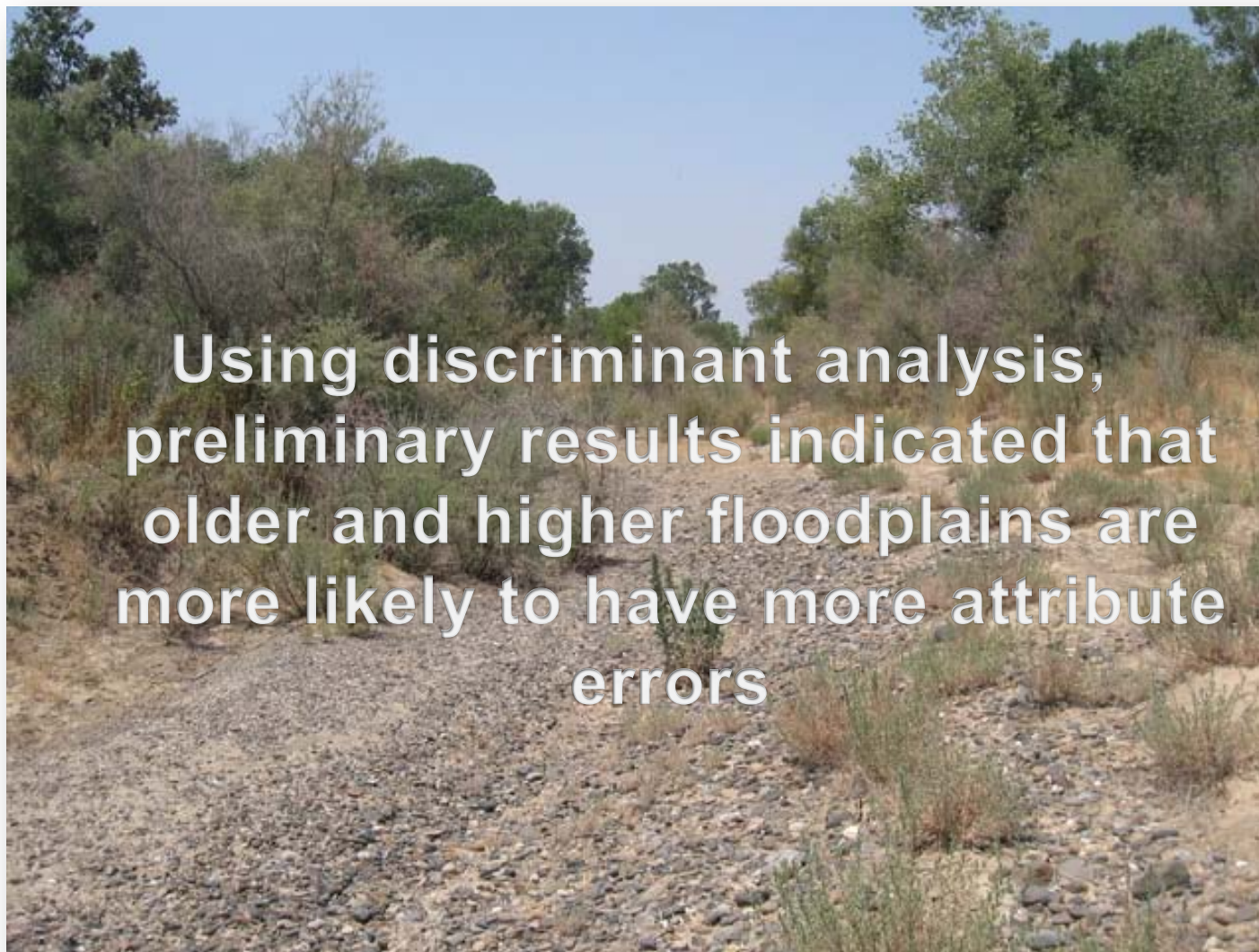
3,300 hectares re-digitized (or 132 blocks)



Found limited differences in polygon size by vegetation class with paired comparisons
Cottonwood Forest (CW) displayed the greatest difference in area



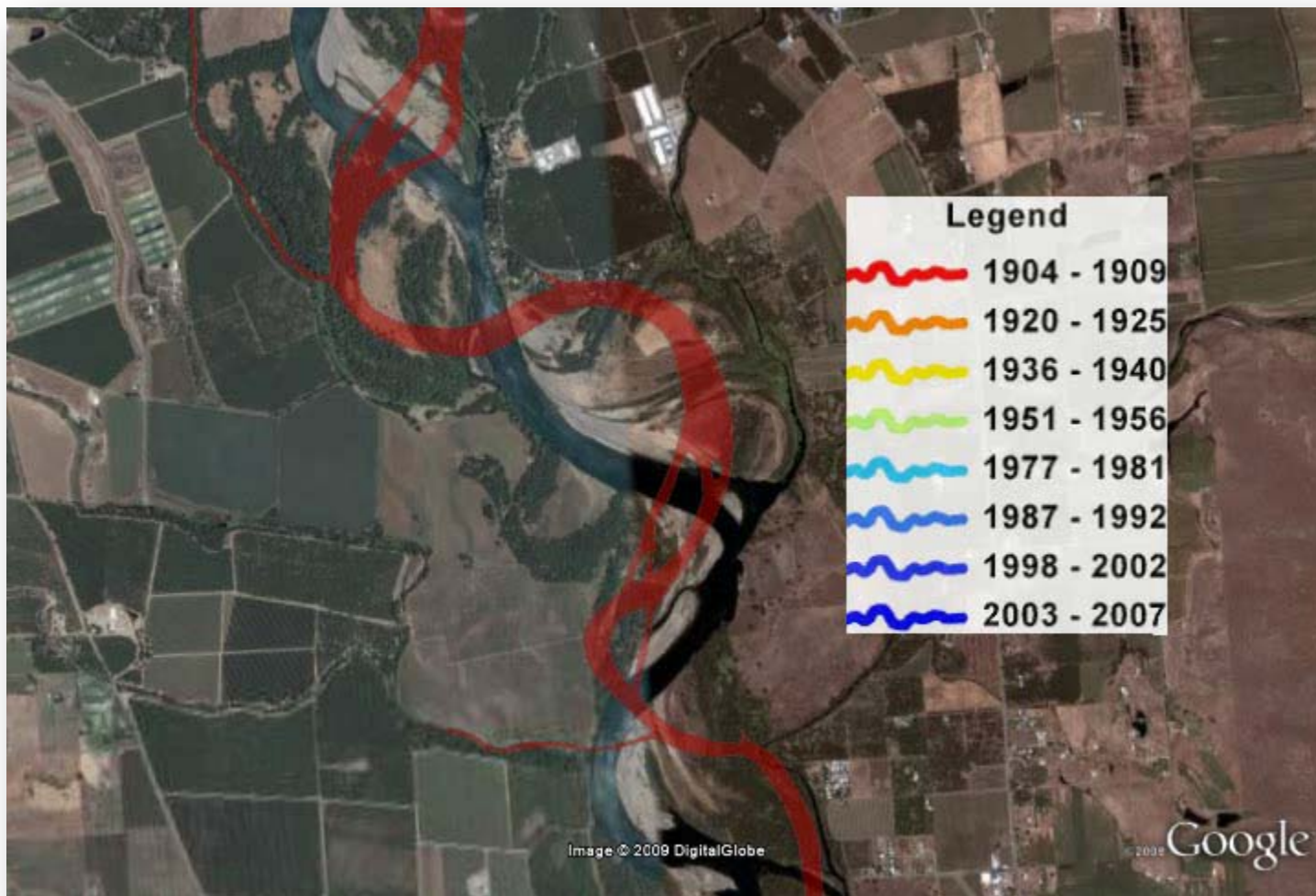
Environmental Variables



Using discriminant analysis, preliminary results indicated that older and higher floodplains are more likely to have more attribute errors



Dynamic Riverscapes





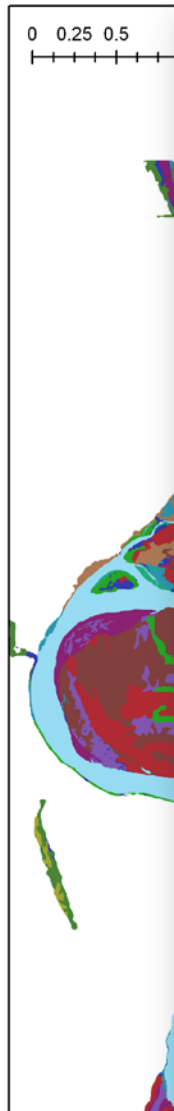
Recursive Partition Analysis



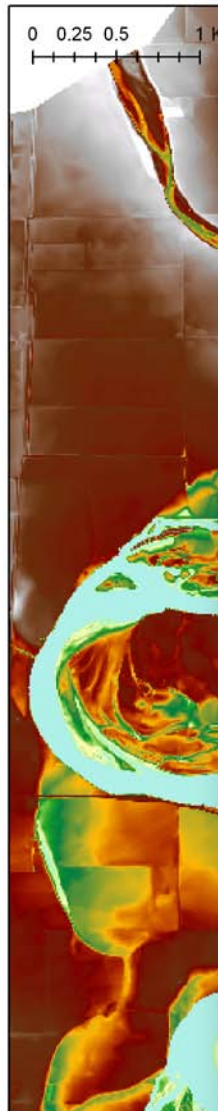
<i>Variable</i>	<i>Description</i>
LnArea	<i>Ln</i> transformed polygon area
Normal PA Ratio	Normalized perimeter to area ratio where /)
Channel Distance	Distance from polygon centroid to main channel
Levee Distance	Distance from polygon centroid to levee
Relative Elevation Range	Range of relative elevation values
Relative Elevation Median	Median relative elevation value
Relative Elevation Minimum	Minimum relative elevation value
Floodplain Age Range	Range of floodplain age values
Floodplain Age Median	Median floodplain age
Floodplain Age Maximum	Maximum floodplain age
Height	Polygon Height Class: 1:<2m, 2:2-6m, 3:6-10m, 4:10-20m, 5:>20m.



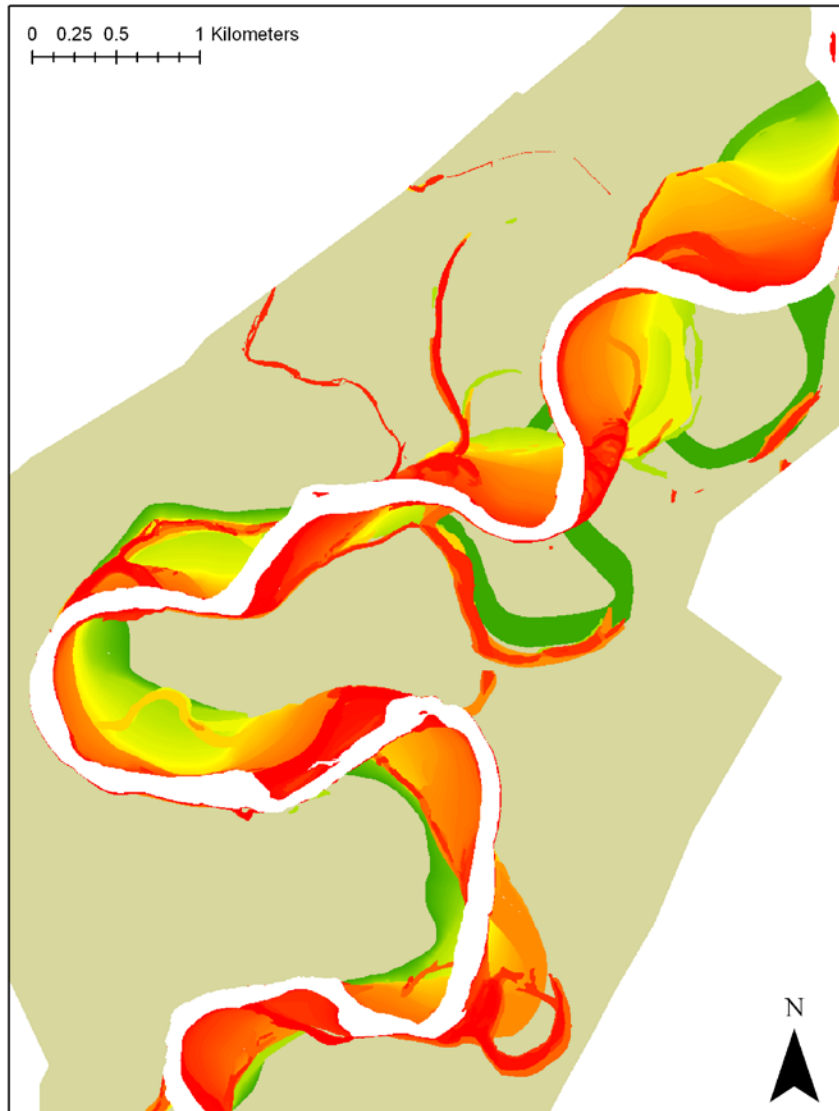
Riverscape Ecology



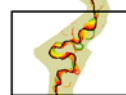
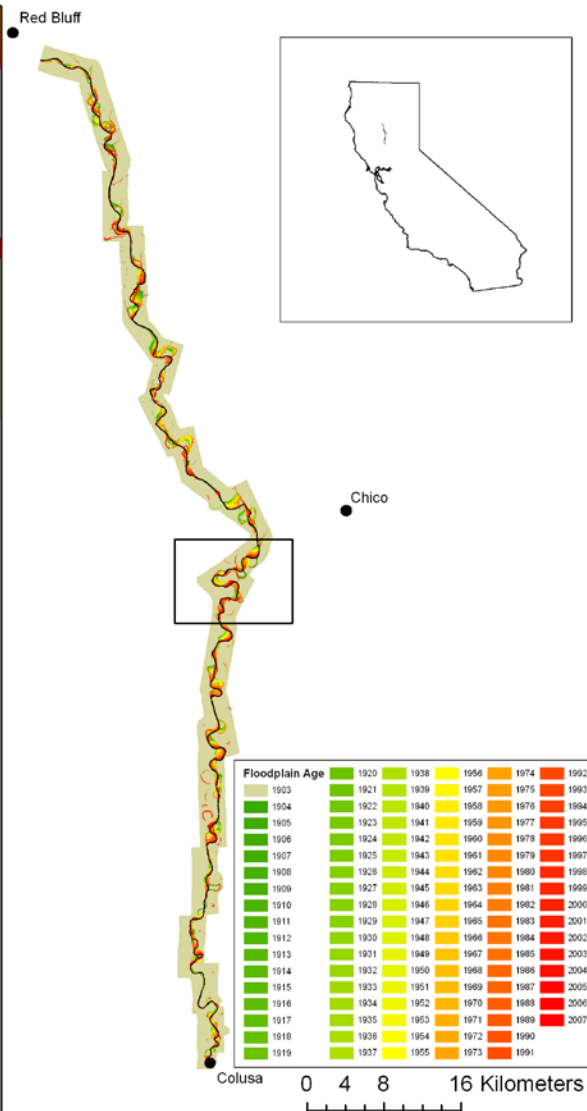
(GIC 2007)



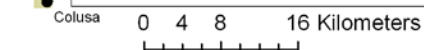
(Greco et al. 2008)



(Fremier In Prep)



Chico



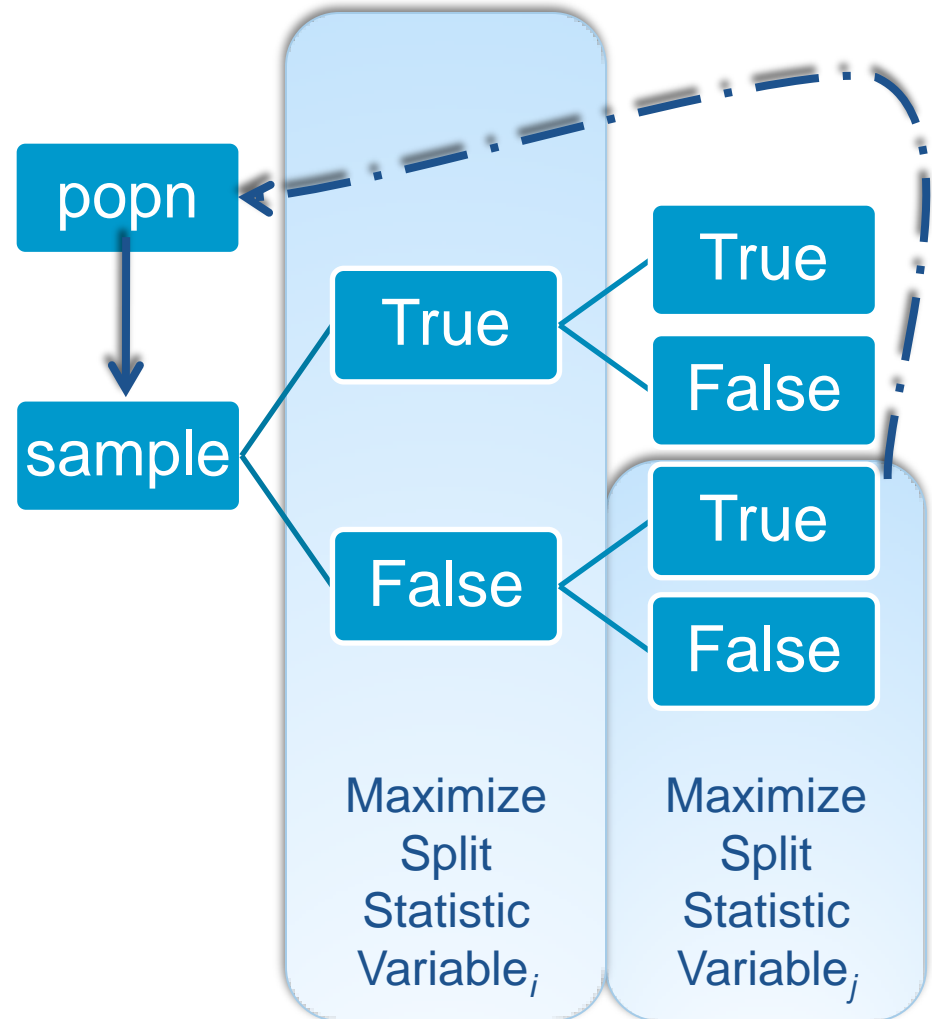


Statistical Partitioning



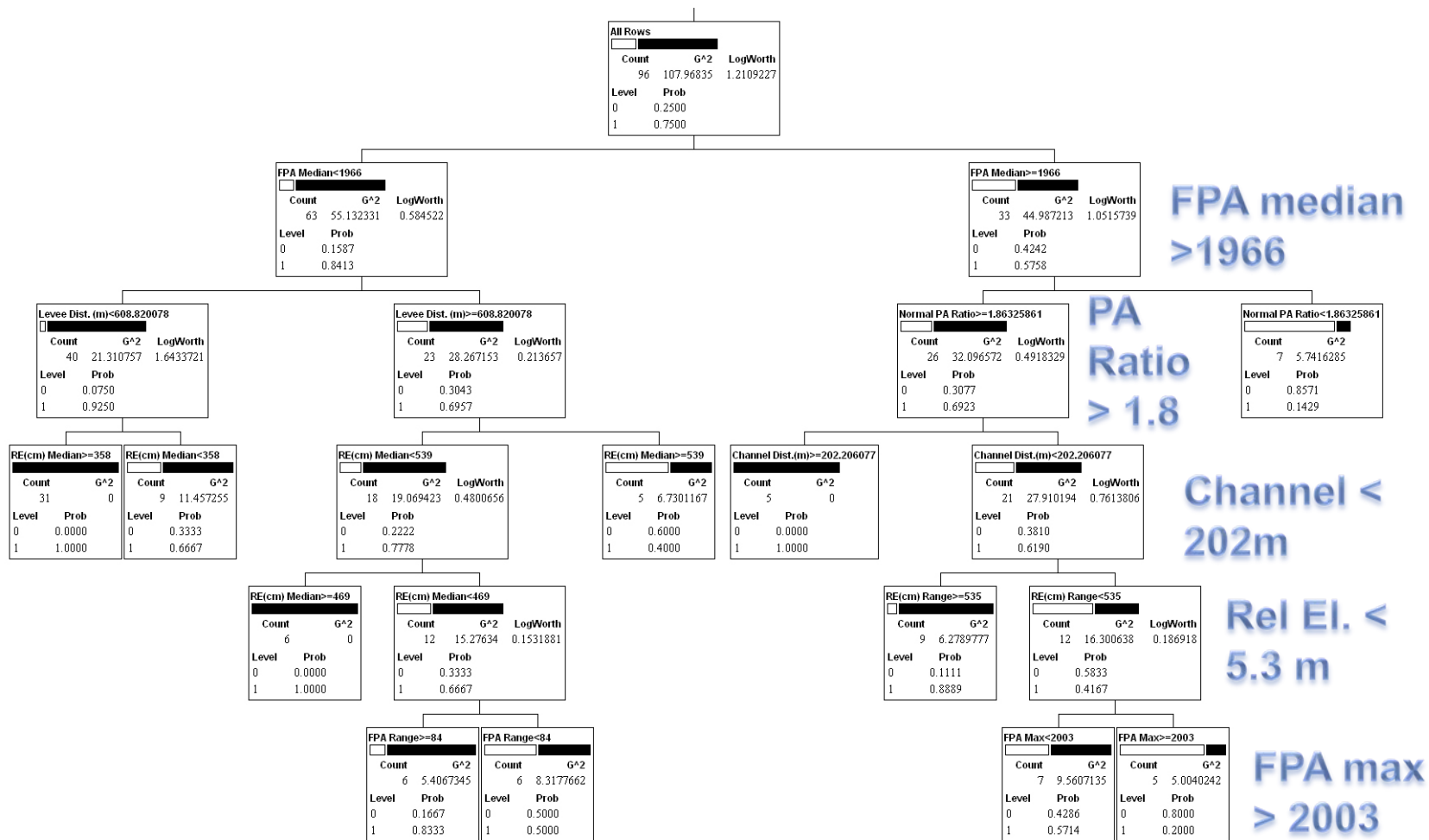
- Recursive partitioning
- Classification & Regression Trees (CART)
- Random Forests

Continuous & Categorical Independent Variables



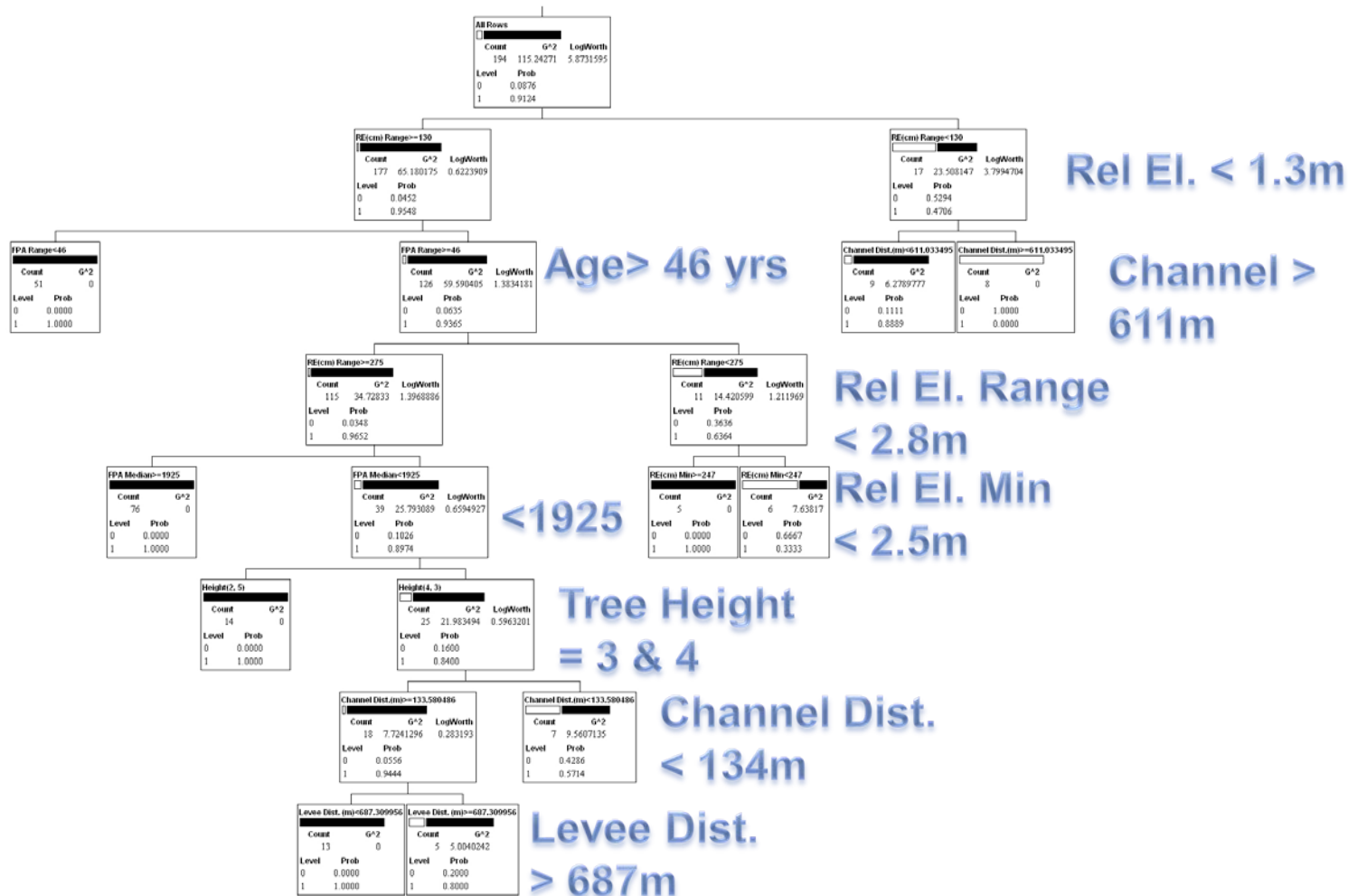


Riparian Scrub (RS)





Cottonwood (CW)



Likelihood of vegetation class correctness as determined by the probability value produced by the recursive partition model.



SRMAP Class Code	n (# polys)	Likely Correct n	Likely Misclass n	Likely Correct ha	Likely Misclass ha	R ²	# of Splits	k-fold
BE BOX ELDER	63	306	77	277.7	61.1	0.56	5	0.50
BW BLACK WALNUT	69	478	101	733.8	158.8	0.46	7	0.27
CA ANNUAL GRASSES	92	444	115	1195.0	370.6	0.65	8	0.54
CW COTTONWOOD	194	723	114	2707.0	327.0	0.75	9	0.69
GB GRAVEL BAR	91	301	66	536.1	81.4	0.63	4	0.49
GW GOODING'S WILLOW	12	6	13	21.1	10.5	0.37	1	0.33
MW MIXED WILLOW	151	593	--	626.8	--	0.60	5	0.49
PG PERENNIAL GRASSES	34	211	52	77.7	26.2	0.32	3	0.13
RS RIPARIAN SCRUB	96	546	213	687.1	288.8	0.46	10	0.27
VO VALLEY OAK	109	559	244	831.0	711.6	0.59	12	0.45

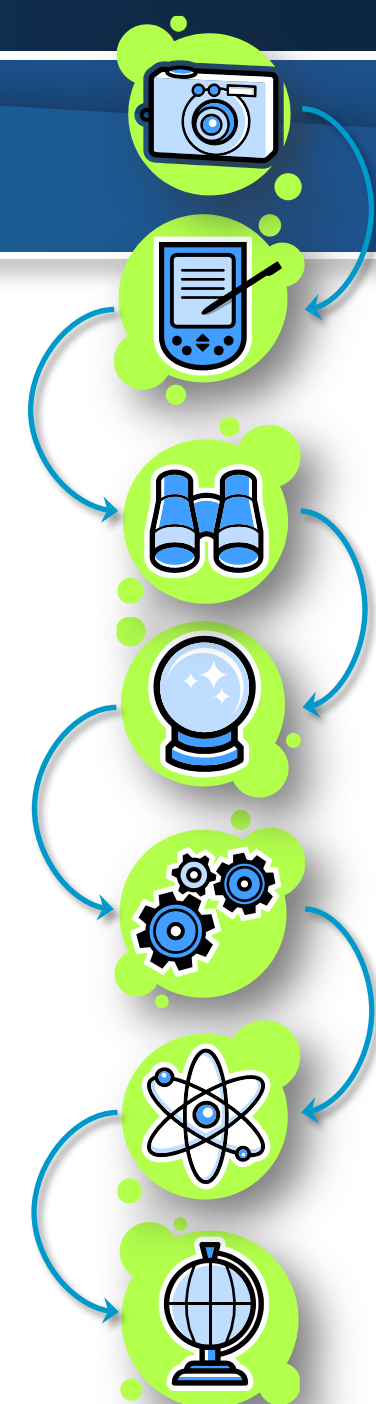
Likely misclassified polygons were identified if they had a >0.5 probability of being incorrectly classified based on the model variables.



Conclusions

- Digitized maps from interpreted aerial imagery will continue to be used, most often because of limitations in resources and expertise, especially in retrospective studies.
- Ancillary datasets, particularly in riverscape ecology, can be used to leverage insights to the spatial context of mapping errors.
- Recursive partitioning is one robust method for crafting type-specific solution sets that combines continuous and categorical spatial data, which can be used to:
 1. ascertain the nature of errors for potential correction (e.g., training sets to fine tune interpretation),
 2. guide map users in interpretation and utility (e.g., removing erroneous polygons from analysis), and
 3. place bounds of confidence around any change detection analyses that are computed from such maps.

or just make more accurate maps!





Acknowledgements

A number of individuals helped make this analysis possible: K. Holl, C. McClain, M. Vaghti, and D. Wood provided field data; E. Larsen and S. Greco provided data layers; L. Calbert, S. Lewis, J. Loyko, and C. Stouthamer provided field assistance; T. Le and M. Jensen helped in the GIS lab; and J. Mount let us borrow his jet boat.



Portions of this project were funded by the CALFED / Bay-Delta Science program's Sacramento River Mapping and Assessment Program (Award# P0620021).

<http://baydelta.ucdavis.edu/srmap/>