

```

1 def FASTMAP(Sn):
2     "Project data on a line between
3     two distant points"
4     z = random.choose(Sn)
5     east = furthest(z, Sn)
6     west = furthest(east, Sn)
7     Sn.poles = (west, east)
8     c = dist(west, east)
9     for one in Sn.members:
10         one.X, one.Y = project(west, east, c, one)
11     return Sn #data members updated with (X,Y)
12
13 def project(west, east, c, x):
14     "Project x onto line east to west"
15     a = dist(x, west)
16     b = dist(x, east)
17     X = (a*a + c*c - b*b)/(2*c), # cosine rule
18     Y = (a*a - X)**0.5 # dist to line b/w poles
19     return X, Y
20
21 def furthest(x, Sn): # furthest from x?
22     out, max = x, 0
23     for y in Sn:
24         d = dist(x, y)
25         if d > max: out, max = y, d
26     return out

```

Figure 1: Splitting data with FASTMAP

```

1 def WHERE4(Sn, min, max):
2     "Recursive cluster quadrants of data"
3     if Sn.length > min:
4         if Sn.length < max:
5             yield Sn
6         else:
7             Xcap, Ycap = mean(Sn.Xs, Sn.Ys)
8             for one in Sn:
9                 S1l += one if one.X < Xcap
10                    and one.Y < Ycap
11                 S1h += one if one.X < Xcap
12                    and one.Y > Ycap
13                 Shl += one if one.X > Xcap
14                    and one.Y < Ycap
15                 Shh += one if one.X > Xcap
16                    and one.Y > Ycap
17             for S in (S1l, S1h, Shl, Shh):
18                 WHERE4(S, min, max)
19     else:
20         yield Sn

```

Figure 2: Recursing Spectrally in quadrants using WHERE4 Algorithm

```

1 def CrossTree(Sn):
2     "Tree of clusters and results
3     from their differences"
4
5     #Calculate (X,Y) of solutions
6     Sn = FASTMAP(Sn)
7     #Accumulate Clusters
8     Cm = [C for C in WHERE4(Sn, min, max)]
9     #Replace data with discretized values
10    Cm = Discretize(Cm)
11    #Score clusters
12    BetterC, WorseC = score(Cm)
13
14    #Build CART Decision Tree
15    Dtree = cart(Cm)
16
17    #Prune same and more leaves
18    for leaf in Dtree:
19        if size(leaf.clusters) > 1:
20            Dtree.prune(leaf)
21    for subtree in Dtree:
22        if size(subtree.uniq_clusters) < 2:
23            Dtree.prune(subtree)
24
25    for Branch in Dtree:
26
27        #Find a better branch for current branch
28        BBranch = Dtree.nearest_best(Branch.C,
29                                     BetterC)
30
31    if BBranch:
32        #Contrast Set(CS) represents diff b/w
33        #limits of Better and Worse Cluster
34        CS = Dtree.differ(WBranch.C,
35                          BBranch.C)
36
37        #Generate new population from model
38        #using Contrast set
39        ContrastSn = model(Size=Branch.C.size←
40                           *10,
41                           ContrastSet)
42
43        q1, q2, q3, q4 += ContrastSn.qs
44
45    else:
46        #Balance distribution
47        q1, q2, q3, q4 += Limits(Branch.C).qs
48
49    #25%, median, 75%, Maximum
50    return q1, q2, q3, q4

```

Figure 3: Generating Deltas between clusters using CrossTree

```

1 def score(Cm):
2     "Divides set of clusters into better
3     and worse based on objectives"
4     better, worse, similar = 0, 0, 0
5     for this in Cm:
6         for other in Cm:
7             for obj in Cm.objectives:
8                 pop1, pop2 = this.pop[obj],
9                     other.pop[obj]
10                #similar check
11                if a12(pop1, pop2) and
12                    bootstrap(pop1, pop2):
13                    similar += 1
14                #different being better or worse
15                elif median(pop1) > median(pop2):
16                    better += 1
17                else:
18                    worse += 1
19                if better > 0 and worse == 0:
20                    scores[this] += 1
21                #top square root of len are better
22                Cm.cut = sorted(scores), scores.len**0.5
23                return Cm[:cut], Cm[cut:]

```

Figure 4: Generating Deltas between clusters using CrossTree

```

1 def discretize(Cm):
2     "Discretizes data into bin and replaces
3     its values with respective bin ids"
4     #find best cut with least entropy
5     #to predict clusters
6     for C in Cm:
7         pairs[d].append(C.id, C.dvalue)
8     for d in decisions:
9         bins[d] = divide(pairs[d])
10
11     def divide(this):
12         lhs, rhs = Counts(),
13             Counts(sym(x) for x in this)
14         for j, x in enumerate(this):
15             maybe = lhs.n/n0*lhs.ent()
16                 + rhs.n/n0*rhs.ent()
17             if maybe < least:
18                 cut, least = j, maybe
19             rhs - sym(x)
20             lhs + sym(x)
21         if cut:
22             return bins += divide(this[:cut])
23                 += divide(this[cut:])
24         else: return bins
25
26     for C, d in Cm, Cm.decisions:
27         C.values[d] = bins.id(C, d)
28
29     return Cm

```

Figure 5: Generating Deltas between clusters using CrossTree

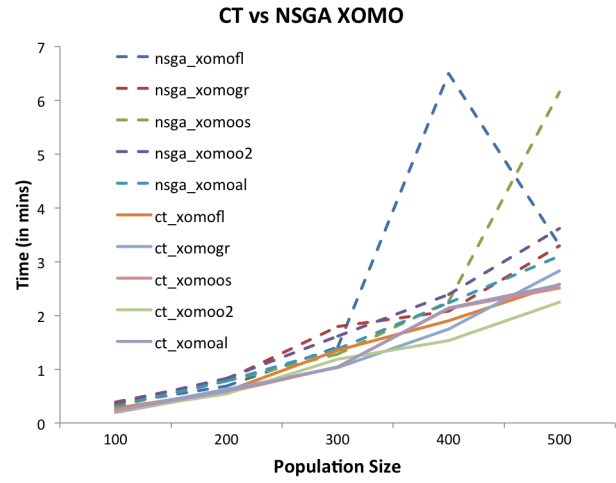


Figure 6: Times for CT and NSGA II on XOMO Model.

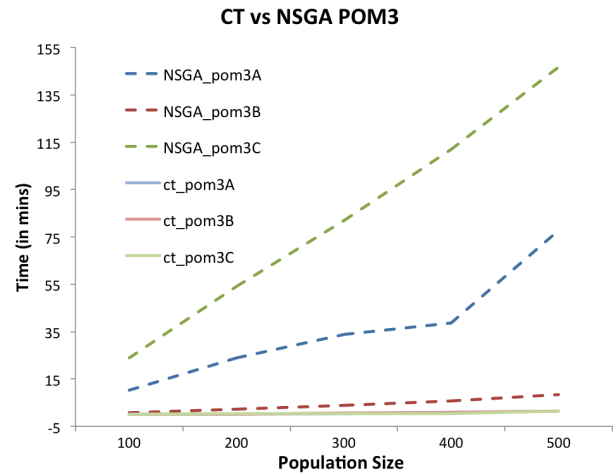


Figure 7: Times for CT and NSGA II on POM Model.

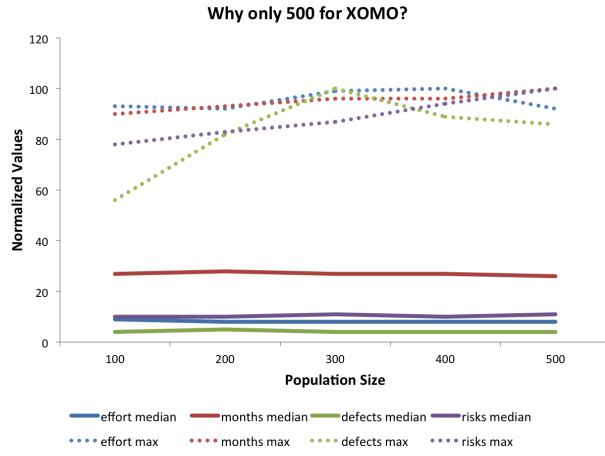


Figure 8: Median and Max of NSGA on XOMO when population size increases from 100 to 500.

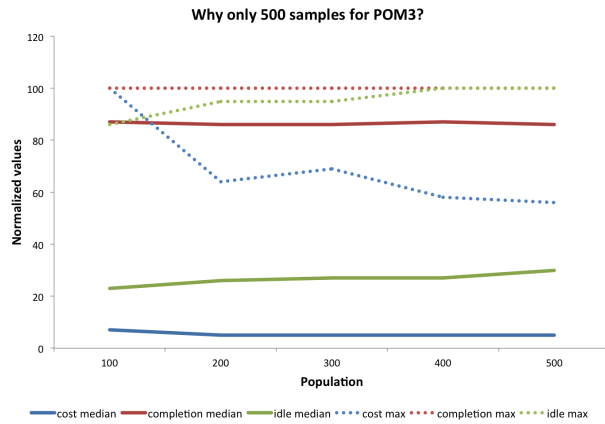


Figure 9: Median and Max of NSGA on POM when population size increases from 100 to 500.

Model: POM3A

Objective	method	median	IQR	
cost	Base Line	32	28	—●—
	CT0	35	25	—●—
	NSGA II	3	0	●
completion	Base Line	85	2	—●—
	CT0	75	0	—●—
	NSGA II	86	1	—●—
idle	Base Line	22	36	—●—
	CT0	17	29	—●—
	NSGA II	31	45	—●—

Model: POM3B

Objective	method	median	IQR	
cost	Base Line	32	27	—●—
	CT0	19	14	—●—
	NSGA II	6	0	●
completion	Base Line	84	2	—●—
	CT0	55	0	—●—
	NSGA II	86	2	—●—
idle	Base Line	29	37	—●—
	CT0	17	27	—●—
	NSGA II	32	47	—●—

Model: POM3C

Objective	method	median	IQR	
cost	Base Line	40	25	—●—
	CT0	31	20	—●—
	NSGA II	10	0	●
completion	Base Line	90	0	—●—
	CT0	63	1	—●—
	NSGA II	90	1	—●—
idle	Base Line	34	33	—●—
	CT0	21	25	—●—
	NSGA II	36	41	—●—

Figure 10: Model: POM3

Model: XOMOAL

Objective	method	median	IQR
effort	Base Line	6	3
	CT0	4	4
	NSGA II	11	12
months	Base Line	20	0
	CT0	19	4
	NSGA II	36	14
defects	Base Line	14	22
	CT0	3	6
	NSGA II	3	5
risks	Base Line	73	17
	CT0	7	6
	NSGA II	5	8

Model: XOMOFI

Objective	method	median	IQR
effort	Base Line	19	15
	CT0	4	5
	NSGA II	11	12
months	Base Line	27	1
	CT0	19	0
	NSGA II	36	13
defects	Base Line	22	29
	CT0	2	3
	NSGA II	5	7
risks	Base Line	87	5
	CT0	2	0
	NSGA II	6	9

Model: XOMOGI

Objective	method	median	IQR
effort	Base Line	7	5
	CT0	3	3
	NSGA II	13	14
months	Base Line	21	0
	CT0	23	2
	NSGA II	39	16
defects	Base Line	16	22
	CT0	1	2
	NSGA II	4	5
risks	Base Line	74	13
	CT0	6	3
	NSGA II	5	8

Figure 11: Model: XOMO

Model: XOMOO2

Objective	method	median	IQR
effort	Base Line	5	0
	CT0	7	1
	NSGA II	13	14
months	Base Line	25	0
	CT0	20	0
	NSGA II	43	20
defects	Base Line	37	24
	CT0	1	0
	NSGA II	5	8
risks	Base Line	79	16
	CT0	13	0
	NSGA II	6	10

Model: XOMOOS

Objective	method	median	IQR
effort	Base Line	3	0
	CT0	8	2
	NSGA II	13	15
months	Base Line	19	0
	CT0	24	2
	NSGA II	41	19
defects	Base Line	7	4
	CT0	1	1
	NSGA II	5	9
risks	Base Line	90	4
	CT0	20	14
	NSGA II	8	13

Figure 12: Model: XOMO

Model: POM3

Model	Value	Cost	Completion	Idle
pom3A	0	30.76	0.12	0.04
pom3B	0	575.69	0.12	0.03
pom3C	0	83.5	0.1	0.08
pom3A	100	1933.55	1.0	0.82
pom3B	100	21916.18	1.0	0.81
pom3C	100	1830.55	1.0	0.69

Model: XOMO

Model	Value	Effort	Months	Defects	Risks
xomool	0	343.25	5.05	2298.1	0.69
xomofl	0	314.32	5.76	1180.89	0.55
xomogr	0	207.79	4.52	969.14	0.72
xomoo2	0	179.9	1.67	1204.4	0.66
xomoos	0	120.06	1.67	1711.01	0.66
xomool	100	6302.63	68.98	103503.64	22.55
xomofl	100	6243.79	67.54	83713.17	20.78
xomogr	100	6177.7	65.31	114610.14	20.41
xomoo2	100	6290.36	63.86	74828.94	18.52
xomoos	100	6739.59	66.83	64776.58	13.55

Figure 13: XOMO and POM3 0100 values

Model: POM3A

Objective	method	median	IQR	
cost	Base Line	14	10	—●—
	CT0	32	29	—●—
	NSGA II	1	0	●
completion	Base Line	86	0	—●—
	CT0	64	0	—●—
	NSGA II	86	0	—●—
idle	Base Line	25	35	—●—
	CT0	13	23	—●—
	NSGA II	32	46	—●—

Model: POM3B

Objective	method	median	IQR	
cost	Base Line	2	1	●
	CT0	21	24	—●—
	NSGA II	0	0	●
completion	Base Line	84	0	—●—
	CT0	64	1	—●—
	NSGA II	86	0	—●—
idle	Base Line	29	39	—●—
	CT0	15	29	—●—
	NSGA II	33	48	—●—

Model: POM3C

Objective	method	median	IQR	
cost	Base Line	40	21	—●—
	CT0	39	23	—●—
	NSGA II	9	0	●
completion	Base Line	90	2	—●—
	CT0	68	0	—●—
	NSGA II	90	1	—●—
idle	Base Line	33	32	—●—
	CT0	20	22	—●—
	NSGA II	35	40	—●—

Figure 14: Model: POM3

	R	Media	
cost	C	13	—●—
	W	45	—●—
	N	2	●
completio	C	82	—●—
	W	82	—●—
	N	88	—●—
idle	C	36	—●—
	W	18	—●—
	N	26	—●—
02	R	media	
effort	C	5	●
	W	7	●
	N	13	—●—
month	C	25	—●—
	W	20	—●—
	N	43	—●—
defect	C	37	—●—
	W	1	●
	N	5	—●—
risks	C	79	—●—
	W	13	—●—
	N	6	—●—

Figure 15: XOMO results (left); POM3 results (right); all results from 20 runs with different random seeds. Big black dots show median values. Horizontal lines show 25th to 75th percentile. All results are normalized 0..100, min..max. Except for POM3's *completion* objective, *smaller* values are *better*. In the "Rx" column, "0,W,N" denotes results from baseline, WICKED, and NSGA-II (respectively).