Supporting Information: Manifold Analysis for High-Dimensional Socio-Environmental Surveys

Charles Dupont and Debraj Roy

Faculty of Science, Informatics Institute, University of Amsterdam,
Science Park 904, 1090, GH, Amsterdam, the Netherlands.
{c.a.dupont,d.roy}@uva.nl

Algorithm Pseudocode

Algorithm 1: t-SNE

Data: $D = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$
Input: dimension d, perplexity, learning rate η, number of steps T, momentum $\alpha(t)$
Result: $Y^{(T)}$, a lower-dimensional representation of the data
begin
  Compute $p_{j|i}$, $p_{i|j}$;
  Compute p_{ij};
  Initialize sample solution $Y^{(0)} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$ with $\mathbf{y}_i \in \mathbb{R}^d$;
  for $t = 1, 2, \ldots, T$ **do**
    Compute q_{ij};
    Compute gradient $\frac{\partial C}{\partial Y} = \frac{\partial}{\partial Y} KL(P||Q)$;
    Update $Y^{(t)} = Y^{(t-1)} + \eta \frac{\partial C}{\partial Y} + \alpha(t)\left(Y^{(t-1)} - Y^{(t-2)}\right)$;
  end
end

Algorithm 2: Diffusion Maps

Data: $D = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$
Input: dimension d, α, ϵ, number of iterations t
Result: Ψ_t, a lower-dimensional representation of the data
begin
  Compute diffusion matrix $L_{i,j} = k(\mathbf{x}_i, \mathbf{x}_j)$;
  Construct diagonal matrix $D_{i,i} = \sum_j L_{i,j}$ from row sums of $L_{i,j}$;
  Normalize the matrix: $L^{(\alpha)} = D^{-\alpha} L D^{-\alpha}$;
  Compute $M = (D^{(\alpha)})^{-1} L^{(\alpha)}$, where $D^{(\alpha)}_{i,i} = \sum_j L^{(\alpha)}_{i,j}$;
  Compute d largest eigenvalues $(\lambda_1^t, \ldots, \lambda_d^t)$ of M^t and corresponding eigenvectors $(\psi_1, \ldots, \psi_d)$;
  Construct embedding $\Psi_t(\mathbf{x}_i) = (\lambda_1^t \psi_1(\mathbf{x}_i), \ldots, \lambda_d^t \psi_d(\mathbf{x}_i))$;
end

Additional Feature Sets

Spatial Characteristics Localization codes corresponding to the district, "upazila", union, and village that each household belongs to are provided in the survey. There are 40 communities in total, each one being defined as a unique combination of the district, upazila, and union codes. Note that upazilas are sub-units of districts, and unions are even smaller administrative units. Each community contains an average of 20.125 households. Each household also possesses one of 7 distinct codes corresponding to different agro-ecological zones. Only agro-ecological and district codes are used for our analysis, although other localization codes are used to bin households into communities. Spatial characteristics are summarized in SI Table 2.

Adaptation Options Households were asked what changes they have implemented in response to climate change. We consider 31 different binary adaptation options, summarized in SI Table 3. Approximately 91% of households adapted in at least one regard, the most implemented adaptation being to change crop varieties (76% adoption rate).

Climatic Shocks Households were also asked whether various weather events adversely impacted the household's activities. There are 13 weather event categories, yielding 13 binary features. Three more specific questions pertaining to weather-related shocks are also included. A summary of these features is presented in SI Table 1. Households were most frequently affected by droughts (56%) and flashfloods (26%). Only 2 households reported being impacted by changing seasons, and just 1 by sea level rise.

Household Characteristics Information was collected about various household characteristics. Continuous features include household size, household head age, number of assets and value, number of lands and value, owned quantities of livestock, as well as expenditure and income. SI Table 4 provides summary statistics of these features. Discrete binary features include: sex of household head (1 for female, 0 for male), whether the primary occupation of the household head is in agriculture (1 if so, 0 if not), the same for their secondary occupation, and whether any member of the household has pursued religious education (1 if so, 0 if not). SI Table 5 provides a summary of these features. Lastly, categorical features include the highest education level amongst household members (SI Table 6), and the type of access to electricity, if any (SI Table 7).

Mutual Information Based Feature Selection

We use a mutual-information based approach to feature selection by extracting features having high mutual information with target features associated with adaptation. The severity level of each constraint to adaptation (no access to

credit, land, input, or money, scarcity of water or labor, no market, no information on climate change and appropriate adaptations, "other") is taken in turn as the target, each feature receives its average mutual information associated with the targets as overall score. The community that households are a part of is the highest-scoring feature, further motivating the choice of forming groups at the community level for FINE. In addition to the community, we retain the top 30 features. SI Table 9 summarizes this set of features.

Discretisation of Continuous Features

We discretise continuous features into bins to facilitate the estimation of probability mass functions using the bayesian blocks dynamic programming method, first introduced by Scargle. This approach seeks an optimal segmentation of a set of continuous data, yielding bins of different sizes, each of which corresponds to a separate uniform distribution. We discretise continuous features into at most 5 bins. A comparison of a raw histogram and Bayesian Blocks results for average monthly household income and expenditure can be found in SI Figure 1.

Feature Set Tables

Table 1: Climatic Shocks

Feature	$\mathbb{P}$(Occurrence)	**N**
Floods	0.08	518
Flashfloods	0.26	518
Drought	0.56	518
Salinity increase	0.05	518
Sea level rise	0.00	518
Frequent rainfall	0.05	518
Temperature/sun intensity increase	0.05	518
Temperature variability	0.01	518
Soil/river erosion	0.03	518
Tornado/very high winds	0.03	518
Cyclone	0.04	518
Seasons changing	0.00	518
Other	0.25	518
Lost home due to river erosion	0.02	552
Major crop loss due to floods	0.12	552
Loss of productive assets due to floods	0.03	552

Table 2: Spatial Characteristics

Feature	**min**	**max**	**N**
District Code	1	31	805
Upazila Code	1	39	805
Union Code	1	40	805
Agro-ecological Code	1	7	805

Table 3: Adaptation Options

Feature	$\mathbb{P}$(Change)	**N**
Decision to adapt	0.91	805
Change crop variety	0.76	729
Change crop type	0.25	729
Change amount of land under production	0.21	729
Implement soil/water management techniques	0.06	729
Change pattern of crop consumption	0.06	729
Mix crop/livestock production at same time	0.01	729
Mix crop/fish farming production at same time	0.05	729
Change field location	0.06	729
Build water harvesting scheme for consumption	0.01	729
Build water harvesting scheme for crops	0.13	729
Build water harvesting scheme for livestock	0.00	729
Build a diversion ditch	0.21	729
Plant trees for shading	0.04	729
Irrigated	0.52	729
Irrigated more	0.61	729
Buy insurance	0.00	729
Change from crop to livestock production	0.01	729
Change from livestock to crop production	0.01	729
Seek off farm employment	0.16	729
Migrate to this place from another place	0.02	729
Set up seed banks/food storage facilities	0.01	729
Change animal breed	0.04	805
Increase number of livestock	0.06	805
Decrease number of livestock	0.04	805
Diversify livestock feeds	0.13	805
Change livestock feeds	0.21	805
Supplement livestock feeds	0.18	805
Change veterinary interventions	0.35	805
Change portfolio of animal species	0.01	805
Move animals to another site	0.00	805

Table 4: Household Characteristics (continuous)

Feature	**mean**	**stdev**	**min**	**max**	**N**
Household size	5.62	2.78	1	23	808
Household head age	47.34	13.91	18.0	97	803
Assets	25.63	25.74	1	473	804
Assets Value	74753.23	103401.43	30	937330	804
Lands	6.02	3.31	1	28	805
Lands Value	1910397.01	2444343.83	4000.0	30245000	805
Cattle Quantity	1.51	1.74	0	9	738
Goat Quantity	0.77	1.56	0	15	738
Pig Quantity	0.02	0.37	0	10	738
Chicken Quantity	11.57	29.85	0	500.0	738
Expenditures (14 day recall)	4520.85	9841.96	227.6	192304.12	805
Monthly income	6925.05	6439.72	0.00	68500.00	804

Table 5: Household Characteristics (binary)

Feature	$\mathbb{P}(1)$	**N**
Household head sex	0.12	803
Primary occupation in agriculture	0.64	803
Secondary occupation in agriculture	0.64	803
Religious education	0.03	808

Table 6: Household Characteristics (highest education level)

No education	**Preschool**	**Years 1-9**	**High School**	**Bachelor's**	**Master's**	**N**
4.6%	0.1%	69.4%	21.9%	2.6%	1.4%	805

Table 7: Household Characteristics (electricity access)

No electricity	**Via national grid**	**Via solar power**	**N**
46.71%	8.82%	44.47%	805

Table 8: Handpicked Features

Feature Description	Median KL Divergence
How likely is it that people who do not participate in community activities will be criticized or sanctioned? (female respondent)	0.349
Monetary loss due to personal shocks	0.329
Severity of constraint to adaptation: no information on climate change and appropriate adaptations	0.316
How likely is it that people who do not participate in community activities will be criticized or sanctioned? (male respondent)	0.306
Severity of constraint to adaptation: shortage of money	0.270
Someone in household works in farming	0.261
Severity of constraint to adaptation: no access to water	0.236
How likely is it that people will cooperate to try to solve water supply problem? (female respondent)	0.208
Monetary loss due to drought	0.207
How likely is it that people will cooperate to try to solve water supply problem? (male respondent)	0.199
Changed planting dates	0.196
Severity of constraint to adaptation: no access to input	0.195
How many people beyond household could you turn to who would be willing to provide money in an emergency? (male respondent)	0.190
Action taken in response to drought	0.185
How many close friends do you have these days? (male respondent)	0.182
# days in past 12 months that someone in household participated in community activities?	0.174
In the past 12 months, how many people with a personal problem have turned to you for assistance? (female respondent)	0.172
Made modifications to cultivated land	0.169
Severity of constraint to adaptation: no access to land	0.167
Total household expenditures (14 day recall)	0.152
How many close friends do you have these days? (female respondent)	0.150
In the past 12 months, how many people with a personal problem have turned to you for assistance? (male respondent)	0.139
In long term emergency, how many people beyond household could you turn to? (female respondent)	0.137
Average monthly income of household	0.136
In long term emergency, how many people beyond household could you turn to? (male respondent)	0.128
Severity of constraint to adaptation: no market	0.125
Changed production for livetock/poultry	0.114
Severity of constraint to adaptation: no access to credit	0.113
Monetary loss due to flashfloods	0.106
Action taken in response to flashfloods	0.105

Economic status of people willing to help you (male respondent)	0.102
Changed production techniques/technology	0.099
Community development activities you worked on with others in community	0.093
Economic status of people willing to help you (female respondent)	0.089
How likely is it that people in the community would get together to help neighbors in unfortunate happenings? (female respondent)	0.083
If you suddenly had to go away for a day or two, could you count on your neighbors to take care of your children? (male respondent)	0.082
How many people beyond household could you turn to who would be willing to provide money in an emergency? (female respondent)	0.081
Community development activities that household worked on with others in community	0.079
Changed fertilizer	0.077
Severity of constraint to adaptation: other	0.070
In the past 12 months have you worked with others in your village/neighborhood to do something for the benefit of the community? (female respondent)	0.067
In the past 12 months have you worked with others in your village/neighborhood to do something for the benefit of the community? (male respondent)	0.063
Someone in household works in wage labor	0.060
Sought off farm employment	0.059
If you suddenly had to go away for a day or two, could you count on your neighbors to take care of your children? (female respondent)	0.058
Changed water harvesting methods	0.057
Severity of constraint to adaptation: shortage of labor	0.055
Administrative activities you worked on with others in community	0.055
Does a group you are part of improve household's livelihood?	0.055
Is a group you are part of important in times of shocks/emergency?	0.054
How likely is it that people in the community would get together to help neighbors in unfortunate happenings? (male respondent)	0.053
Administrative activities that household worked on with others in community	0.049
Someone in household is self employed	0.048
Monetary loss due to floods	0.044
Someone in household is a trader	0.034
Someone in household is a salaried worker	0.025
Does this group help in improving living?	0.020
Does this group address livelihood diversification to your household?	0.019
Changed crop consumption	0.013
Action taken in response to floods	0.011
Does a group you are part of benefit the community?	0.005
Does this group help in risk management?	0.000
Action taken in response to salinity increase	0.000

Set up communal seed banks/food storage facilities	0.000
Does this group help in climate change adaptation responses?	0.000
Monetary loss due to cyclone	0.000
Action taken in response to frequent rainfall	0.000
Does this group provide climate/weather information?	0.000
Action taken in response to tornato/very high winds	0.000
Action taken in response to sea level rise	0.000
Action taken in response to soil/river erosion	0.000
Action taken in response to temperature variability	0.000
Monetary loss due to tornato/very high winds	0.000
Other activities you worked on with others in community	0.000
Monetary loss due to frequent rainfall	0.000
Action taken in response to temperature/sun intensity increase	0.000
Does this group provide information on crop prices/market opportunities?	0.000
Changed something else	0.000
Other activities that household worked on with others in community	0.000
Is a group you are part of important for other reasons?	0.000
Monetary loss due to temperature/sun intensity increase	0.000
Is a group you are part of for enjoyment/recreation?	0.000
Action taken in response to cyclone	0.000
Monetary loss due to temperature variability	0.000
Action taken in response to seasons changing	0.000
Monetary loss due to soil/river erosion	0.000
Does this group provide/link in agricultural inputs or advice?	0.000
Is a group you are part of important for various information?	0.000
Social activities you worked on with others in community	0.000
Social activities that household worked on with others in community	0.000
Someone in household works in livestock/poultry service	0.000
Monetary loss due to salinity increase	0.000
Migrated	0.000
Someone in household works in production	0.000
Monetary loss due to sea level rise	0.000

Table 9: Top Mutual Information Features

Feature Description	**Average Mutual Information**
Community code	1.333
Agro-ecological code	0.423
Severity of constraint to adaptation: no access to input	0.403
Severity of constraint to adaptation: shortage of money	0.380
Success rate of joint community petitions	0.324
Total weight of "other" food category consumed	0.299
Severity of constraint to adaptation: no access to land	0.271
Expenditure on "drink" food category	0.236
Severity of constraint to adaptation: no info. on climate change and appropriate adaptations	0.234
Miscellaneous expenditures	0.234
What is the main source of drinking water?	0.231
Total weight of "drink" food category consumed	0.229
Made modifications to cultivated land	0.228
Changed production techniques/technology	0.228
Severity of constraint to adaptation: no market	0.226
Severity of constraint to adaptation: no access to credit	0.217
Monetary value of positive economic events	0.199
Severity of constraint to adaptation: scarcity of water	0.196
Monetary value of agricultural equipment owned	0.183
Frequency of contact with extension agents	0.182
Whether joining a particular group helped with improving living	0.168
Expenditure on "cereals" food category	0.167
Expenditure on family events	0.166
Expenditure on "edible oil" food category	0.166
How many close friends you you have these days? (male respondent)	0.160
Whether joining a particular group provides information on crop prices/market opportunities	0.160
Expenditure on "other" food category	0.159
Does information you receive from extension agents meet your needs?	0.156
Expenditure on "fruit" food category	0.156
In past 12 months have you attended a village/neighborhood council meeting, public hearing, or public discussion group? (male respondent)	0.155
In past 12 months, how many people with a personal problem turned to you for assistance? (female respondent)	0.152

Additional Figures

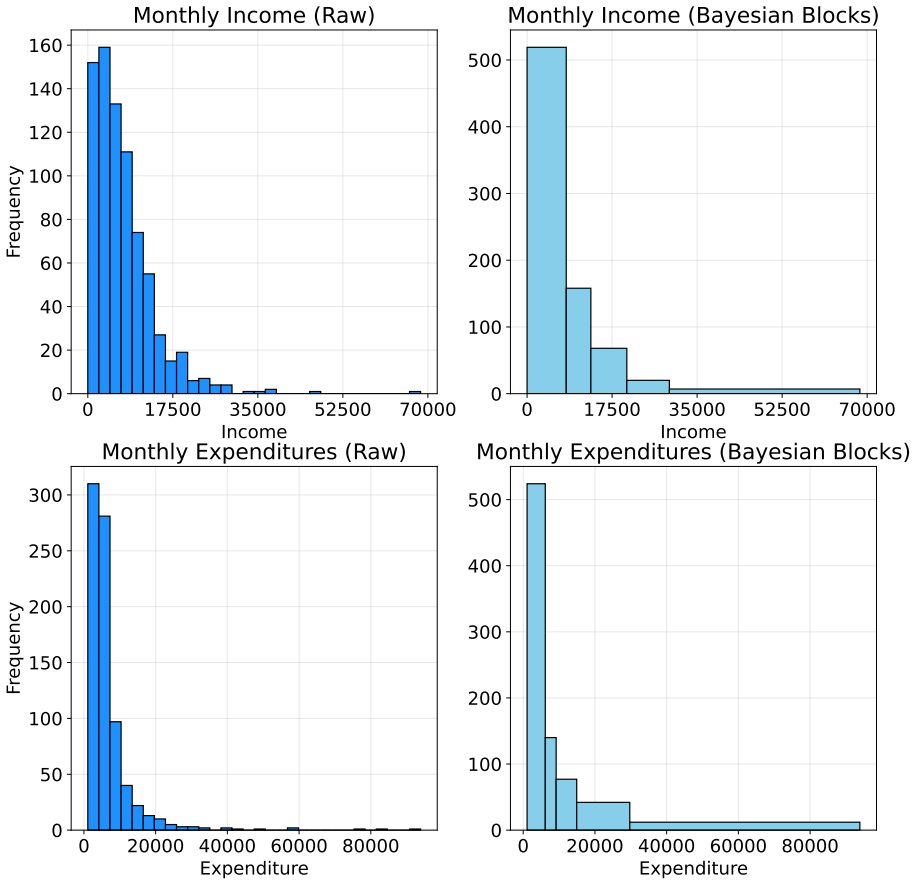

Fig. 1: Comparison of raw data (30 bins of equal width) with bayesian block binning for monthly income and expenditure. Due to both distributions being "fat-tailed", we notice that the bayesian blocks method constructs an especially wide bin for high income and expenditure values.

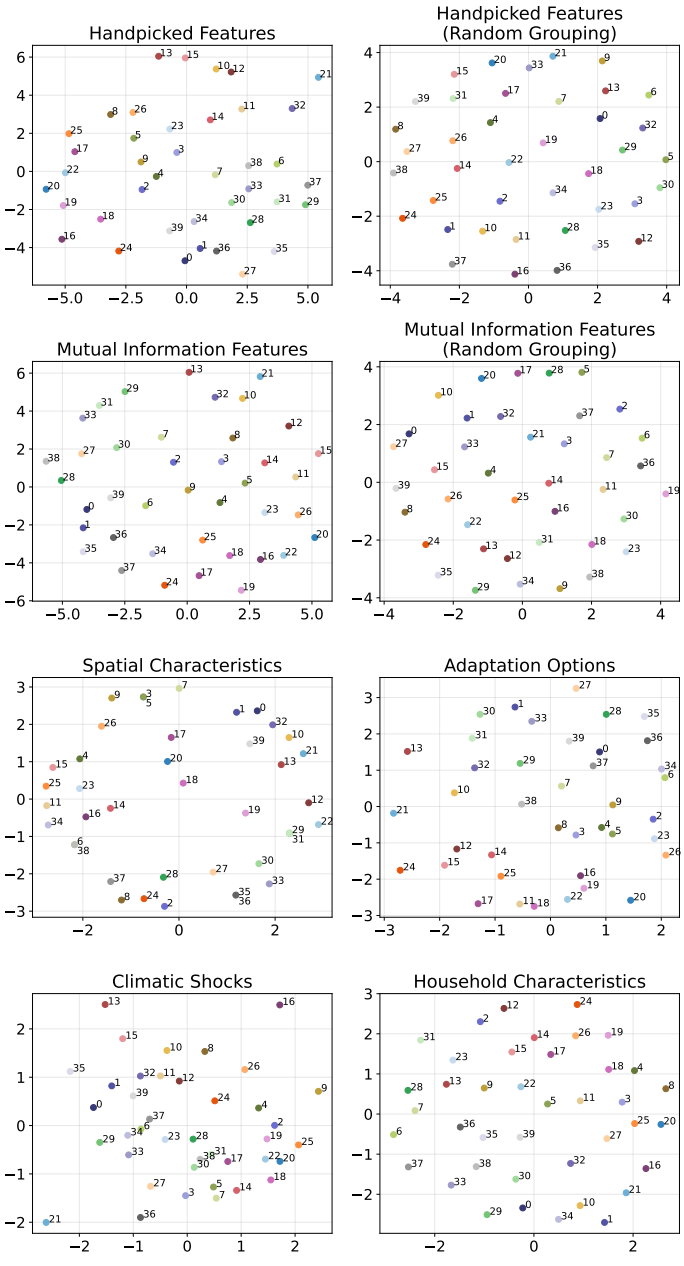

Fig. 2: FINE embeddings for various feature sets. Aside from two randomized grouping experiments, groups are defined according to community membership.

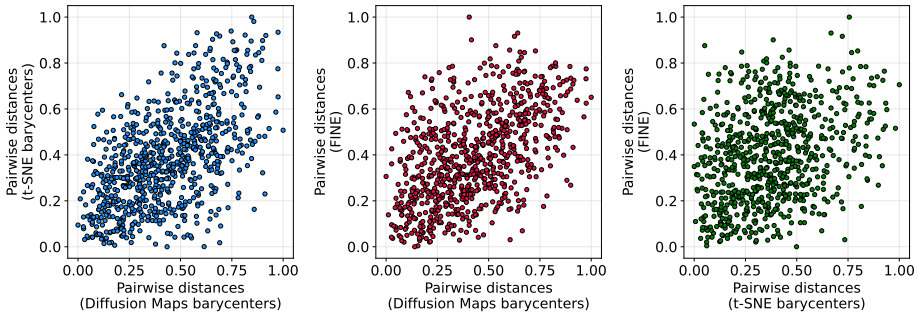

Fig. 3: Comparison of pairwise distances between community coordinates for each pair of algorithms using the handpicked feature set.