Granular Knowledge Representation and Inference using Labels and Label Expressions

Jonathan Lawry

j.lawry@bris.ac.uk

Dept. Engineering Mathematics,

University of Bristol,

Bristol, BS8 1TR, UK

Overview

- Information granules and granular models.
- Representing imprecise labels.
- A prototype interpretation of labels.
- Imprecise probabilities from imprecise descriptions.
- Applications: Case Studies.
- Conclusion.

Information Granules

- Fundamental to human communication is the ability to effectively describe the continuous domain of sensory perception in terms of a finite set of description labels.
- *Granular modelling* permits us to process and transmit information efficiently at an appropriate level of detail.
- *Information Granule* [Zadeh] A granule is a clump of objects (points) which are drawn together by indistinguishability, similarity, proximity and functionality.
- *Information Granule* [Lawry] An information granule is a characterisation of the relationship between a discrete label or expression and elements of the underlying (often continuous) domain which it describes.

Granular Models

- *Granular Modelling* aims to develop formal representations on information granules and embed these into computational models (intelligent systems).
- *Granular models* (Zadeh) are high-level (often rule-based) models which incorporate concepts as represented by information granules.
- Mechanisms for representing and processing uncertainty and linguistic vagueness are fundamental.
- Model transparency and accuracy are dual goals.
- Traceability of decision processes is vital in many applications.
- Key problem areas are: learning, fusion and reasoning.

Deciding What to Say...

- Describing the world requires us to make decisions about what is the best choice of words when referring to objects and instances, in order to convey the information we intend.
- Suppose you are witness to a robbery, how should you describe the robber so that police on patrol in the streets will have the best chance of spotting him?
- You will have certain labels that can be applied, for example *tall, short, medium, fat, thin, blonde, etc*, but which are more appropriate (according to convention)?
- We contend that this is fundamentally an epistemic problem (Williamson).

The Epistemic Stance

- *Within a population of communicating agents, individuals assume the existence of a set of labelling conventions for the population governing what linguistic labels and expression can be appropriately used to describe particular instances.*
- Making an assertion to describe an object or instance x involves making a decision as to what labels can be appropriately used to describe x.
- An individual's knowledge of labelling conventions is partial and uncertain.
- This will result in uncertainty about the appropriateness of labels to describe instance x.
- In accordance with De Finetti it is reasonable to model this uncertainty using subjective probabilities.

Measures of Appropriateness

- As a (big) simplification...
- Assume that there is a finite set of labels $LA = \{L_1, \ldots, L_n\}$ for describing elements of the universe Ω.
- LE is the set of expressions generated from LA through recursive application of the connectives $\wedge, \vee$ and $\neg$.
- $LA = \{red, blue, green, yellow \ldots\}$
  $LE = \{red\&blue, not\ yellow, green\ or\ not\ yellow \ldots\}$
- For $\theta \in LE$, $x \in \Omega$, $\mu_\theta(x) =$ the subjective probability that θ is appropriate to describe x.

Mass Functions

- $\mathcal{D}_x$ is the complete set of labels appropriate to describe x.
- $\mathcal{D}_x = \{red, pink\}$ means that both *red* and *pink* can be appropriately used to describe x, and no other labels are appropriate.
- $m_x : 2^{LA} \to [0,1]$ is a probability mass function on subsets of labels.
- For $F \subseteq LA$ $m_x(F)$ is the subjective probability that $\mathcal{D}_x = F$.
- The mass function m_x and the appropriateness measure μ are strongly related...
- $\mu_\theta(x)$ is the sum of m_x over those values for $\mathcal{D}_x$ consistent with θ.

General Relationships

'x is θ' requires that $\mathcal{D}_x \in \lambda(\theta)$ where $\forall \theta, \varphi \in LE$

- $\forall L \in LA\ \lambda(L) = \{F \subseteq LA : L \in F\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$
- $\lambda(\neg\theta) = \lambda(\theta)^c$

This results in the following equation relating m_x as μ:

- $\mu_\theta(x) = P(\mathcal{D}_x \in \lambda(\theta)) = \sum_{S \in \lambda(\theta)} m_x(S)$
- $\lambda(red \wedge \neg pink) = \{F \subseteq LA : red \in F, pink \notin F\}$
- $\mu_{red \wedge \neg pink}(x) = \sum_{F : red \in F, pink \notin F} m_x(F)$

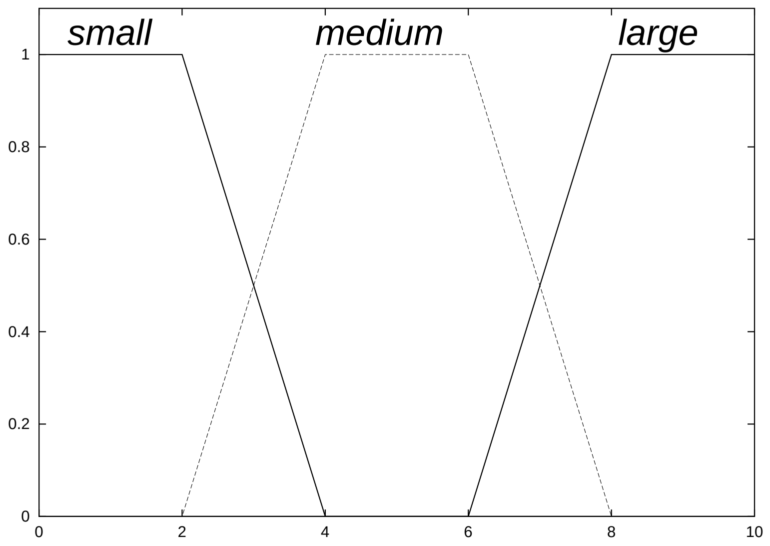
Some Links ...

- *...to probability* For a fixed x $\mu_\theta(x)$ forms a probability measure on LE as θ varies.
- If $\models \theta$ then $\forall x,\ \mu_\theta(x) = 1$
- If $\theta \equiv \varphi$ then $\forall x,\ \mu_\theta(x) = \mu_\varphi(x)$
- $\forall \theta, \forall x,\ \mu_{\neg\theta}(x) = 1 - \mu_\theta(x)$
- If $\models \neg(\theta \wedge \varphi)$ then $\mu_{\theta\vee\varphi}(x) = \mu_\theta(x) + \mu_\varphi(x)$
- *...to DS Theory* For conjunctions and disjunctions of labels appropriateness measure can be interpreted as Commonality and Plausibility functions.
- $\mu_{L_1\wedge\ldots\wedge L_k}(x) = \sum_{F:\{L_1,\ldots,L_k\}\subseteq F} m_x(F) = Q(\{L_1,\ldots,L_k\}|x)$
- $\mu_{L_1\vee\ldots\vee L_k}(x) = \sum_{F:F\cap\{L_1,\ldots,L_k\}\neq\emptyset} m_x(F) = Pl(\{L_1,\ldots,L_k\}|x)$

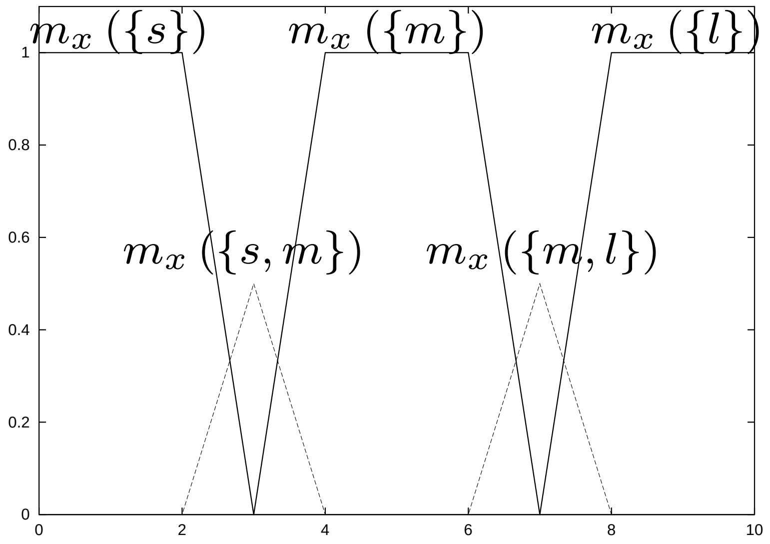
Assuming Consonance

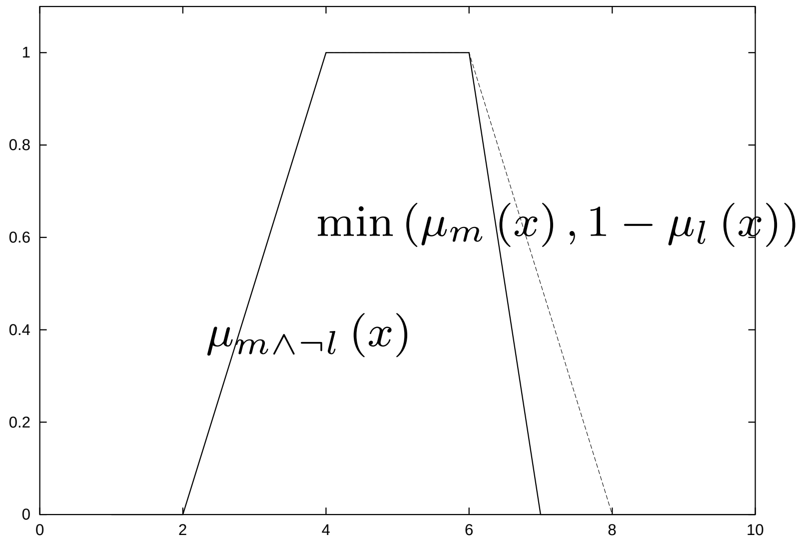
- *Focal Sets* $\mathcal{F}_x = \{F \subseteq LA : m_x(F) > 0\}$.
- *Consonance* For $F, F' \in \mathcal{F}_x$ then either $F \subseteq F'$ or $F' \subseteq F$.
- *Ordering Labels* Assume that for each $x \in \Omega$ an agent first identifies a total ordering on the appropriateness of labels. They then evaluate their belief values m_x about which labels are appropriate to describe x in such a way so as to be consistent with this ordering.
- $LE^{\wedge,\vee} =$ expressions only involving the connectives $\wedge$ or $\vee$.
- *Under Consonance* $\forall \theta, \varphi \in LE^{\wedge,\vee}$, $\forall x \in \Omega$ it holds that $\mu_{\theta\wedge\varphi}(x) = \min(\mu_\theta(x), \mu_\varphi(x))$ and $\mu_{\theta\vee\varphi}(x) = \max(\mu_\theta(x), \mu_\varphi(x))$

Consonance and Functionality



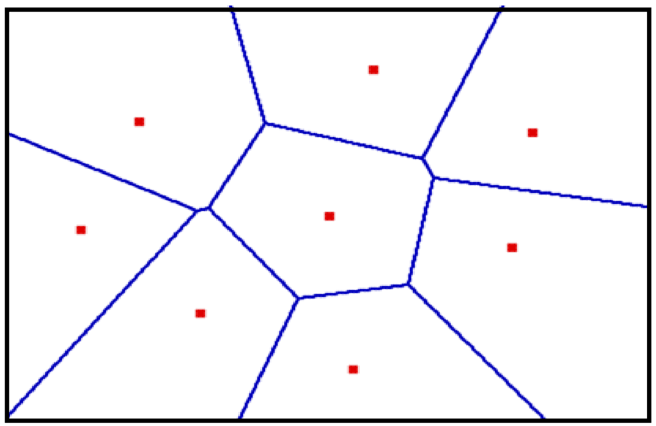
$\mu_{L_1}(x), \ldots, \mu_{L_n}(x)$ ordered so that
$\mu_{L_i}(x) \geq \mu_{L_{i+1}}(x)$ for $i = 1, \ldots, n-1$
$m_x(\{L_1, \ldots, L_n\}) = \mu_{L_n}(x)$
$m_x(\{L_1, \ldots, L_i\}) = \mu_{L_i}(x) - \mu_{L_{i+1}}(x)$
and $m_x(\emptyset) = 1 - \mu_{L_1}(x)$



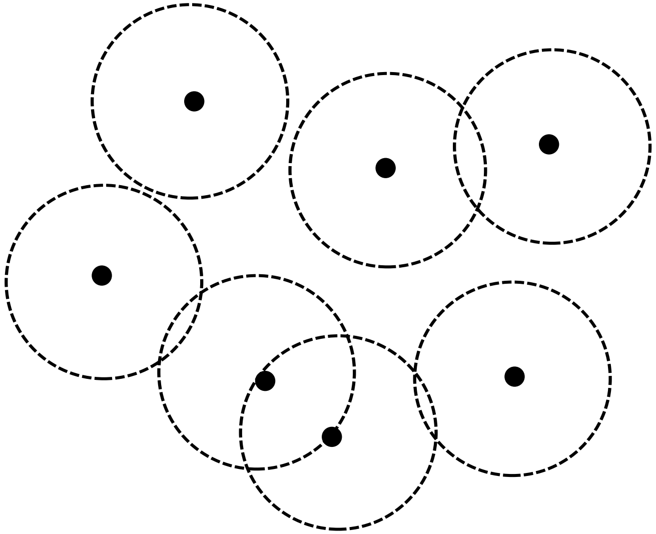


Prototype Theory

- The membership of elements in a concept is determined by their similarity to certain prototypical cases (Rosch 1973).
- Prototypes may be actual exemplars of the concept (case-based reasoning) or abstractions or aggregations (clustering).

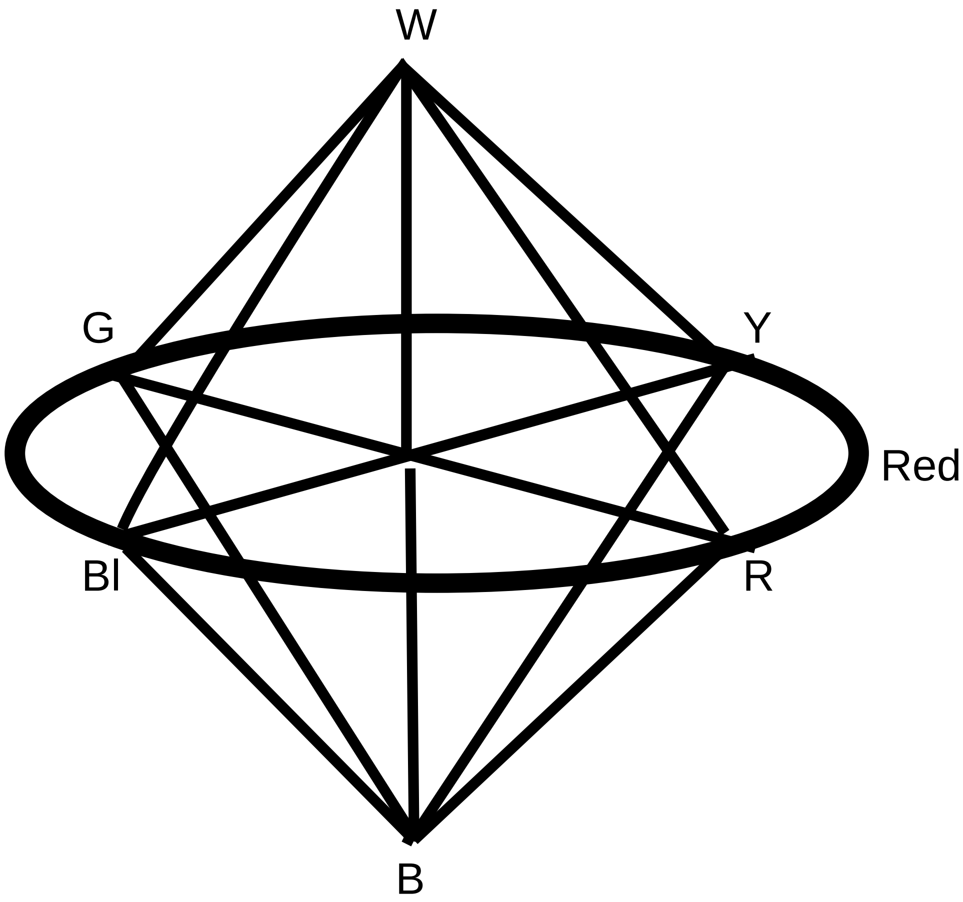


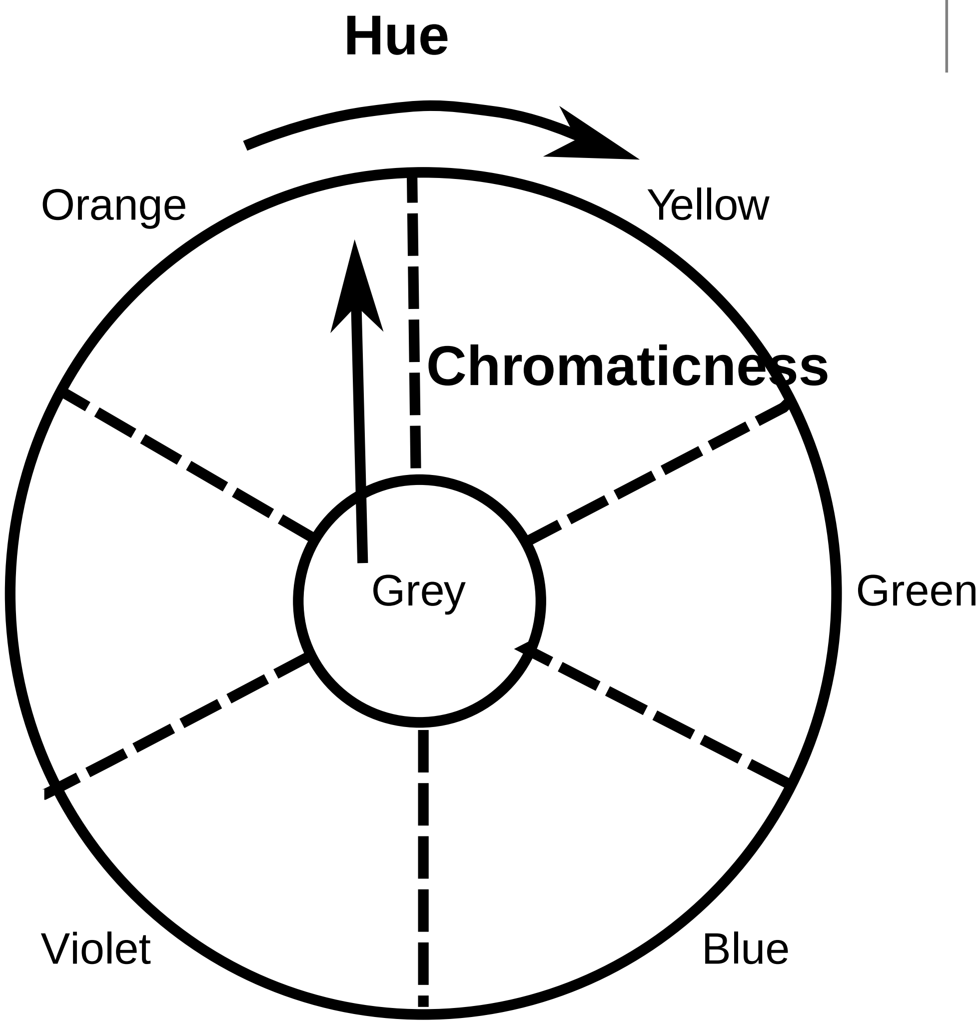
Voronoi Model



Neighbourhood Model

Conceptual Spaces: NCS Colour Spindle

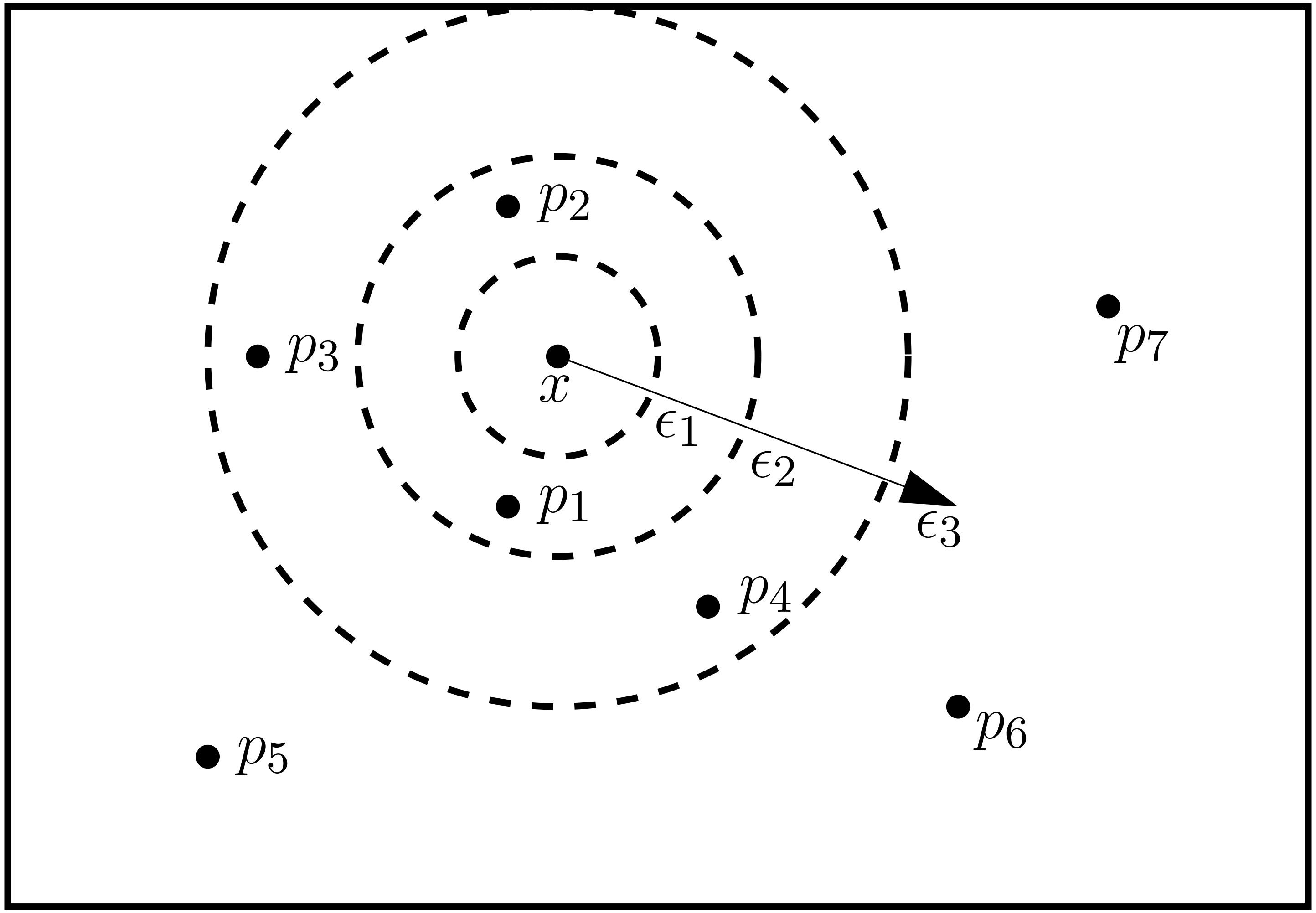




A Prototype Interpretation

- Let $d : \Omega^2 \rightarrow [0, \infty)$ be a distance function satisfying $d(x, x) = 0$ and $d(x, y) = d(y, x)$.
- For $S, T \subseteq \Omega$ let $d(T, S) = \inf\{d(x, y) : x \in S, y \in T\}$.
- For $L_i \in \Omega$ let $P_i \subseteq \Omega$ be a set of prototypical elements for L_i.
- Let ϵ be a random variable into $[0, \infty)$ with density function δ.
- L_i is appropriate to describe x iff $d(x, P_i) \leq \epsilon$.
- $\mathcal{D}_x^\epsilon = \{L_i : d(x, P_i) \leq \epsilon\}$.
- $\forall F \subseteq \Omega \; m_x(F) = \delta(\{\epsilon : \mathcal{D}_x^\epsilon = F\})$
- Naturally satisfies consonance.

Generating $\mathcal{D}_x^\epsilon$



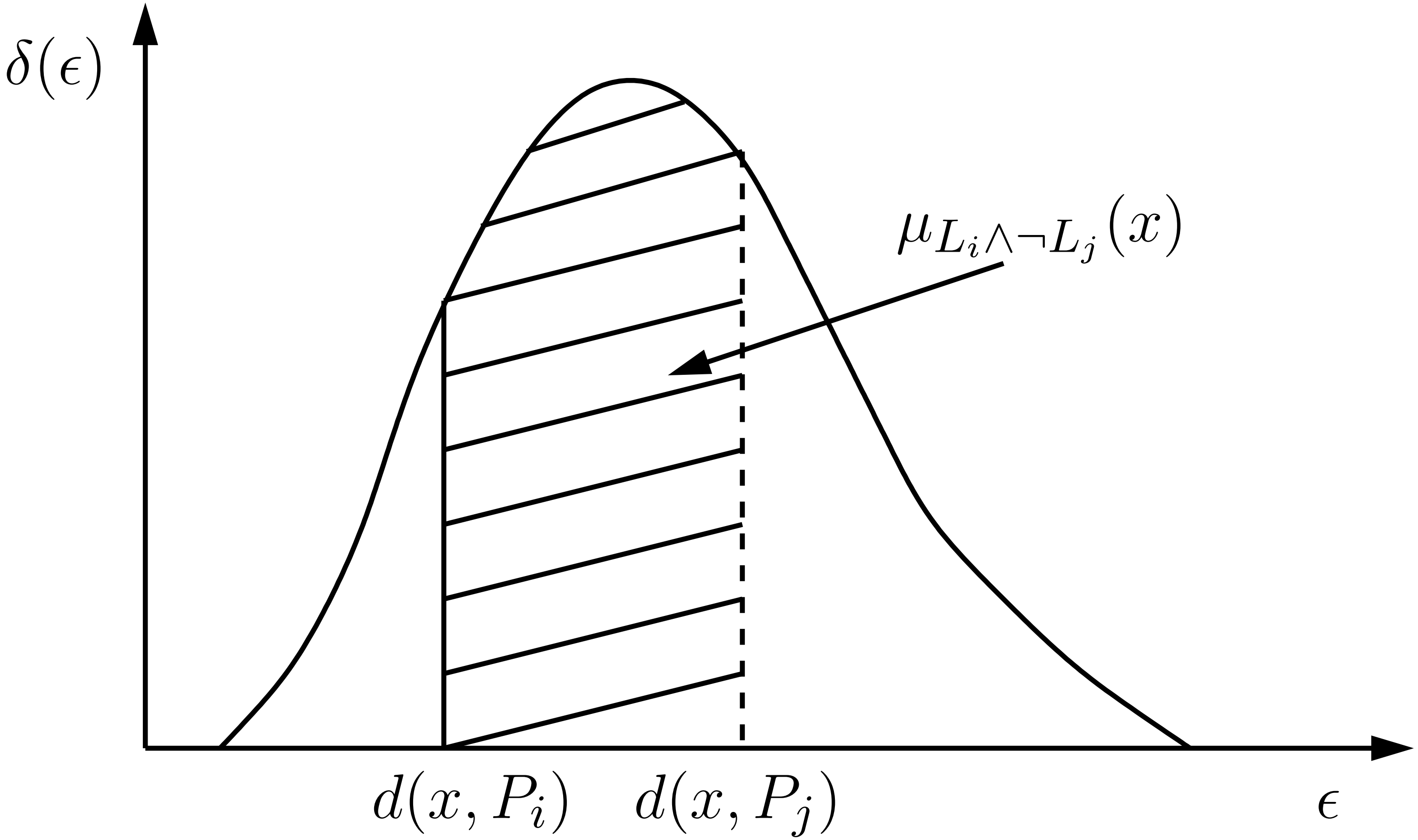
Identifying $\mathcal{D}_x^\epsilon$; $\mathcal{D}_x^{\epsilon 1} = \emptyset$, $\mathcal{D}_x^{\epsilon 2} = \{L_1, L_2\}$, $\mathcal{D}_x^{\epsilon 3} = \{L_1, L_2, L_3, L_4\}$

Appropriateness as Subsets of ϵ

- $\forall x \in \Omega$ and $\forall \theta \in LE$, $I(\theta, x) \subseteq [0, \infty)$ is defined recursively by:
- $\forall L_i \in LA \; I(L_i, x) = [d(x, P_i), \infty)$
- $\forall \theta \in LE \; I(\neg\theta, x) = I(\theta, x)^c$
- $\forall \theta, \varphi \in LE \; I(\theta \vee \varphi, x) = I(\theta, x) \cup I(\varphi, x)$
- $\forall \theta, \varphi \in LE \; I(\theta \wedge \varphi, x) = I(\theta, x) \cap I(\varphi, x)$
- Theorem: $\forall \theta \in LE, \forall x \in \Omega \; I(\theta, x) = \{\epsilon : \mathcal{D}_x^\epsilon \in \lambda(\theta)\}$
- Corollary: $\forall \theta \in LE, \forall x \in \Omega \; \mu_\theta(x) = \delta(I(\theta, x))$

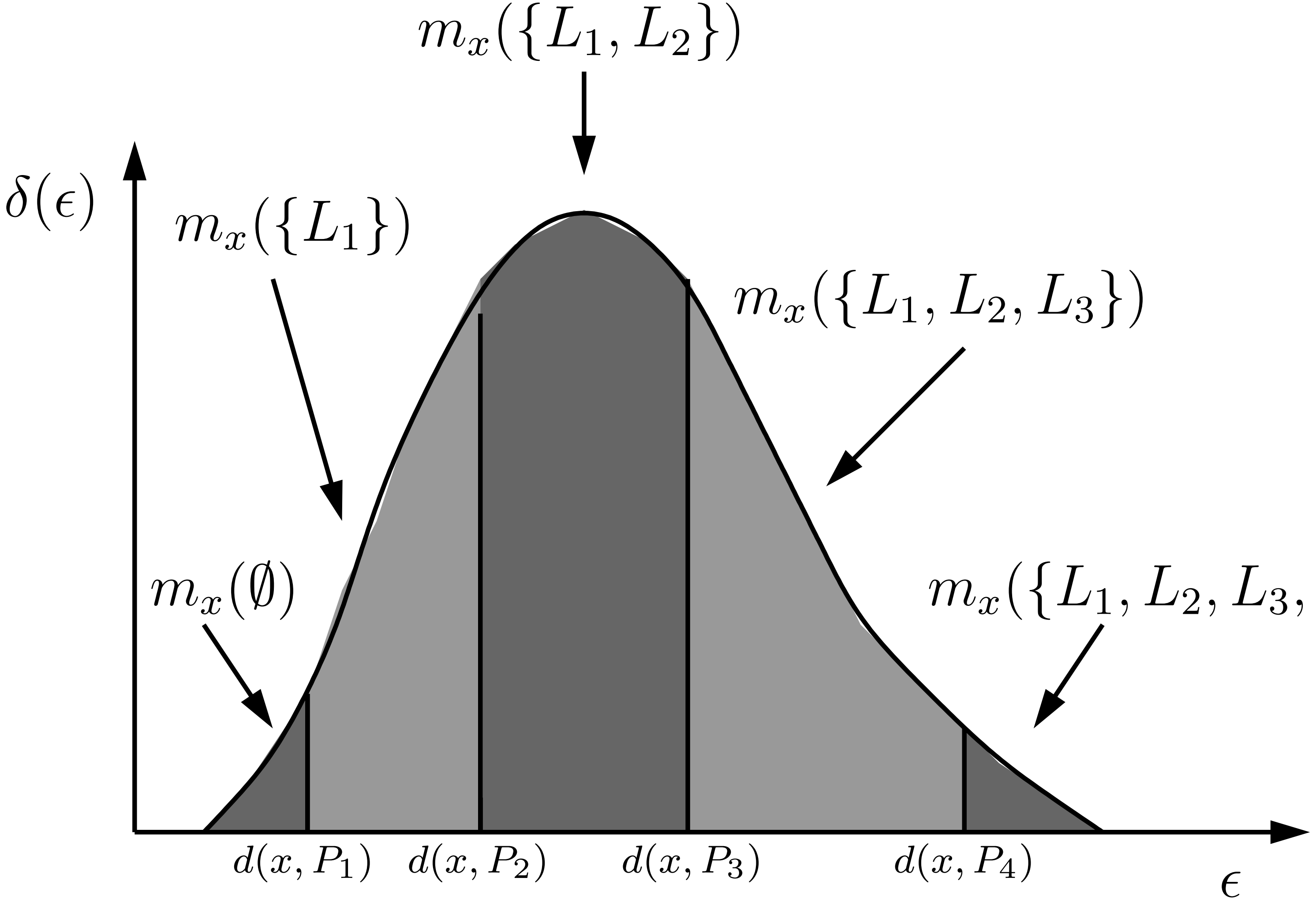
Area under δ: Appropriateness

$$I(L_i \wedge \neg L_j, x) = [d(x, P_i), d(x, P_j))$$



Area under δ: Mass

$$m_x(F) = \delta([\max\{d(x,P_i) : L_i \in F\}, \min\{d(x,P_i) : L_i \notin F\}))$$

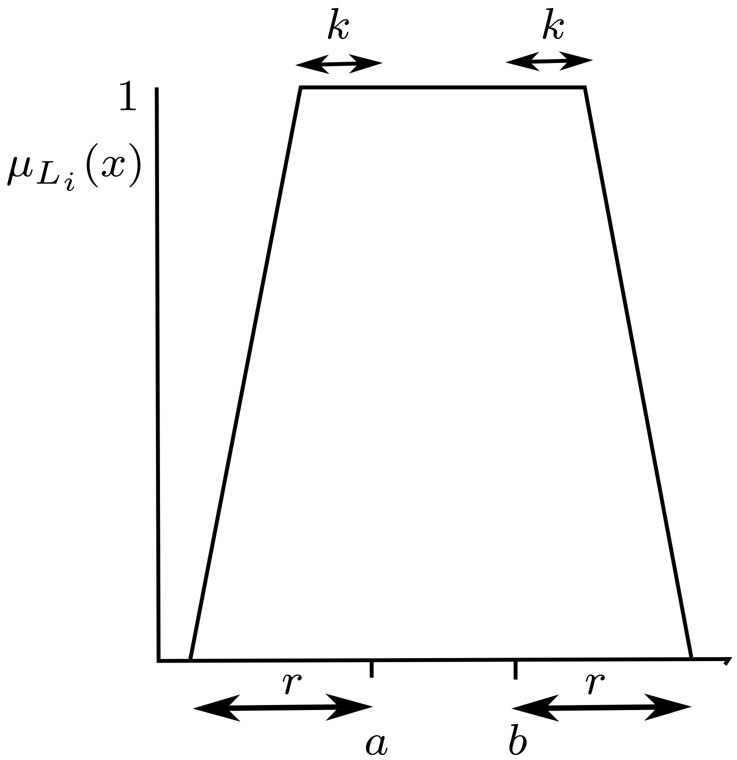


Neighbourhood Representation

- Let $\mathcal{N}^{\epsilon}_{L_i} = \{x \in \Omega : d(x, P_i) \leq \epsilon\}$, more generally $\mathcal{N}^{\epsilon}_{\theta} = \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\}$
- Satisfies $\mathcal{N}^{\epsilon}_{\theta \wedge \varphi} = \mathcal{N}^{\epsilon}_{\theta} \cap \mathcal{N}^{\epsilon}_{\varphi}$, $\mathcal{N}^{\epsilon}_{\theta \vee \varphi} = \mathcal{N}^{\epsilon}_{\theta} \cup \mathcal{N}^{\epsilon}_{\varphi}$, $\mathcal{N}^{\epsilon}_{\neg\theta} = (\mathcal{N}^{\epsilon}_{\theta})^{c}$.
- For $\theta \in LE^{\wedge,\vee}$ $\mathcal{N}^{\epsilon}_{\theta}$ is nested, otherwise not in general.
- Alternative characterisation: $\mu_{\theta}(x) = \delta(\{\epsilon : x \in \mathcal{N}^{\epsilon}_{\theta}\})$
- Labels represent sets of points *sufficiently* similar to prototypes (Information Granules).

Example (Uniform δ)

- Let $\Omega = \mathbb{R}$ and $d(x, y) = ||x - y||$. Let $L_i = about\ [a, b]$ where $a \leq b$, so that $P_i = [a, b]$.
- $\mathcal{N}^{\epsilon}_{L_i} = [a - \epsilon, b + \epsilon]$
- Let δ be the uniform distribution on $[k, r]$ for $0 \leq k < r$:



Random Sets

- A random set is simply a set valued variable.
- Random sets are known to be a unifying concept in uncertainty theory e.g. DS theory, possibility theory, fuzzy theory all have random set interpretations.
- $\forall x \in \Omega\ \mathcal{D}_x$ is a random set into 2^{LA} (i.e. taking sets of labels as values).
- $\forall \theta \in LE\ \mathcal{N}_\theta^\epsilon$ is a random set into 2^Ω (i.e. taking subsets of the underlying universe Ω as values).
- *Single Point Coverage* (of $\mathcal{D}_x$) $P(L_i \in \mathcal{D}_x) = \mu_{L_i}(x)$
- (of $\mathcal{N}_\theta^\epsilon$) $P(x \in \mathcal{N}_\theta^\epsilon) = \mu_\theta(x)$

Imprecise Probabilities

- Suppose an agent receives the information 'x is θ' what information does this tell them about x?
- According to our prototype model, the agent can infer $x \in \mathcal{N}_\theta^\epsilon$
- From this they can define upper and lower probabilities: For $S \subseteq \Omega$
- $\overline{P}(S|\theta) = P(\mathcal{N}_\theta^\epsilon \cap S \neq \emptyset) = \delta_\theta(\{\epsilon : \mathcal{N}_\theta^\epsilon \cap S \neq \emptyset\})$
- $\underline{P}(S|\theta) = P(\mathcal{N}_\theta^\epsilon \subseteq S) = \delta_\theta(\{\epsilon : \mathcal{N}_\theta^\epsilon \subseteq S\})$
- Single point coverage: $spc(x|\theta) = P(x \in \mathcal{N}_\theta^\epsilon) = \delta_\theta(\{\epsilon : x \in \mathcal{N}_\theta^\epsilon\}) \propto \mu_\theta(x)$

Possibility Measures

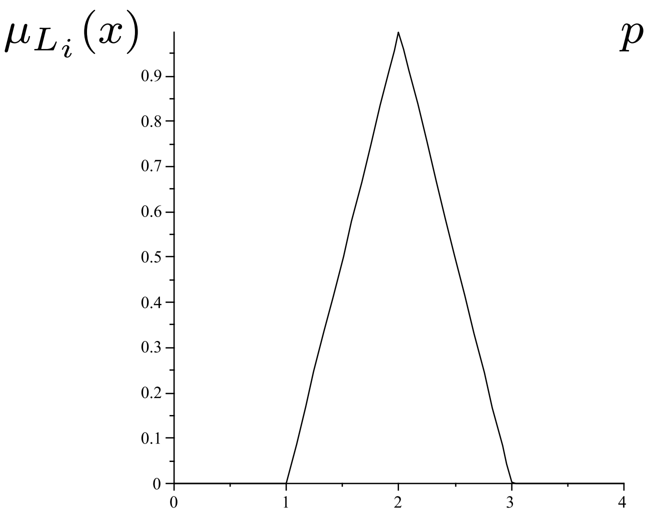
- If $\theta \in LE^{\wedge,\vee}$ then for $0 \leq \epsilon \leq \epsilon'$ we have that $\mathcal{N}_\theta^\epsilon \subseteq \mathcal{N}_\theta^{\epsilon'}$.
- In this case the resulting upper and lower probabilities are possibility and necessity measures (Dubois, Prade).
- $\overline{P}(S|\theta) = Pos(S|\theta) = \sup\{\mu_\theta(x) : x \in S\}$
- $\underline{P}(S|\theta) = Nec(S|\theta) = 1 - \sup\{\mu_\theta(x) : x \in S^c\}$
- With properties...
- $Pos(S \cup T|\theta) = \max(Pos(S|\theta), Pos(T|\theta))$
- $Nec(S \cap T|\theta) = \min(Nec(S|\theta), Nec(T|\theta))$
- Hence $spc(x|\theta) = \mu_\theta(x)$ is a possibility distribution (Zadeh)

Given a Prior...

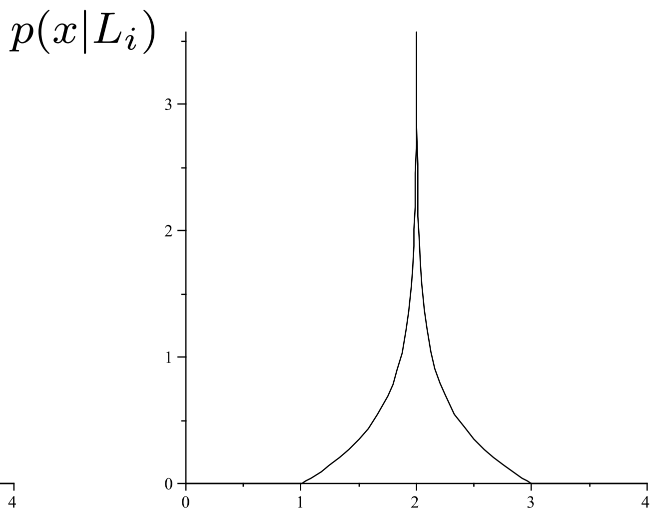
- Suppose that agent also has knowledge in the form of a prior distribution on Ω.
- Given 'x is θ' they should determine the posterior $p(x|\mathcal{N}_\theta^\epsilon)$.
- But since ϵ is uncertain this results in *second order probabilities*.
- If precise probabilities are requires one solution would be to take an expected value:
- $p(x|\theta) = \int_0^\infty p(x|\mathcal{N}_\theta^\epsilon) d\epsilon$.
- This satisfies $p(S|\theta) \in [\underline{P}(S|\theta), \overline{P}(S|\theta)]$

Example (Fuzzy Numbers)

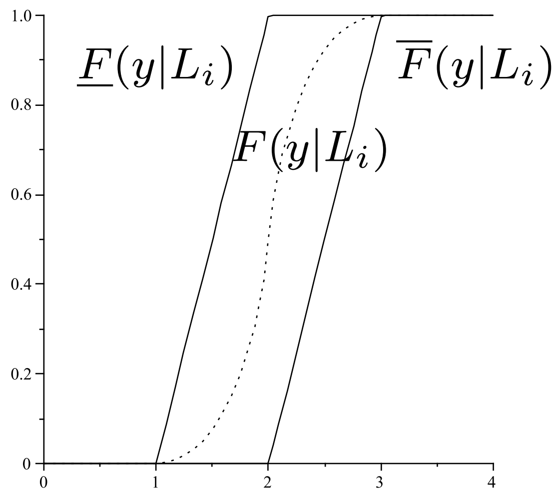
- Let $\Omega = \mathbb{R}$ and $d(x, y) = ||x - y||$. Let $L_i = about\ 2$ so that $P_i = \{2\}$. In this case $\mathcal{N}^{\epsilon}_{L_i} = [2 - \epsilon, 2 + \epsilon]$
- Let δ be the uniform distribution on $[0, 1]$.
- Let the prior p on Ω correspond to the uniform distribution on $[0, 10]$.



Appropriateness



Posterior

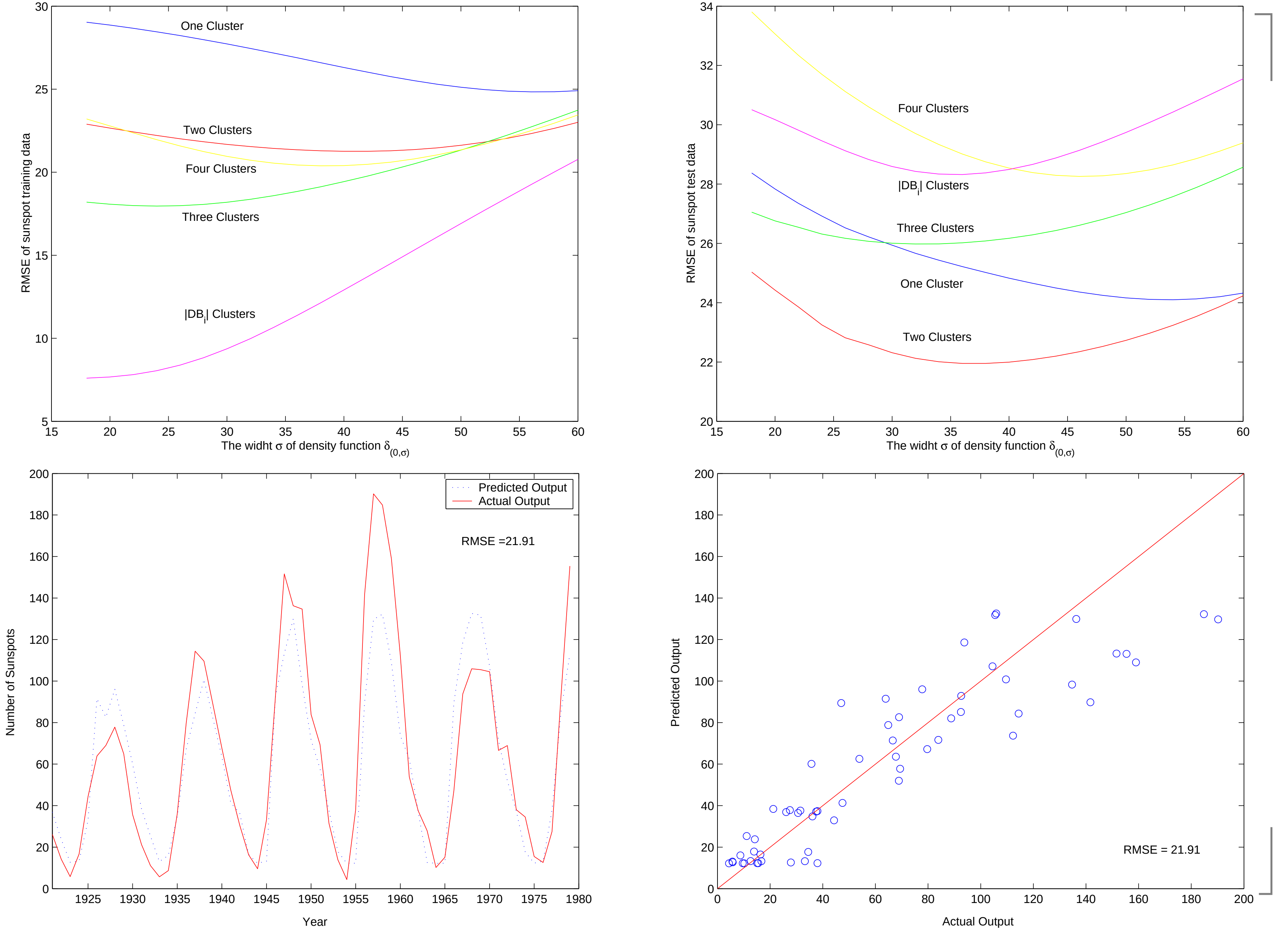


Envelope

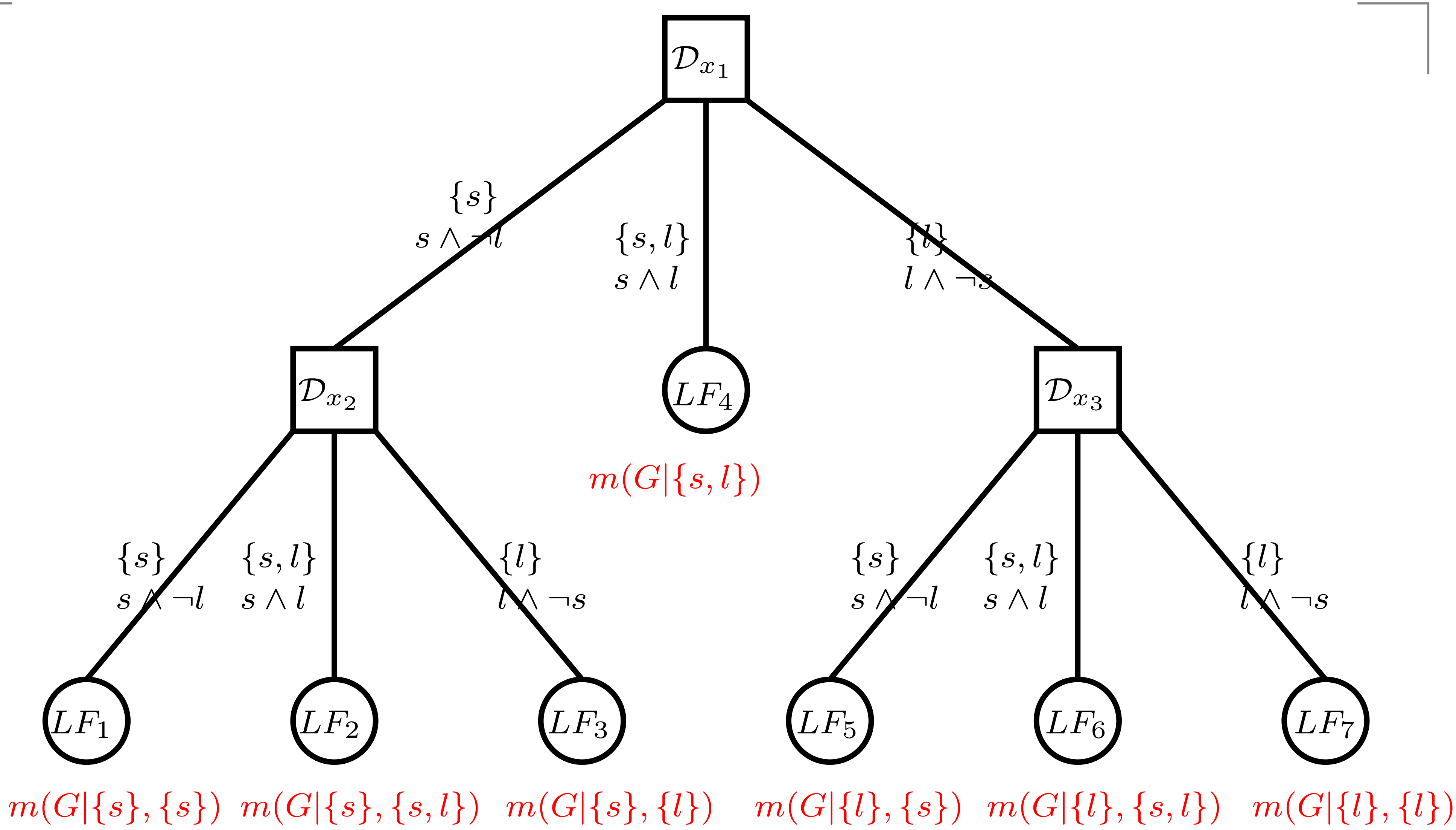
Prototype-Based Rule-Learning

- Joint work with Yongchuan Tang from Zhejiang University.
- Given a database of pairs $(\vec{x}, y)$ where $y = f(\vec{x})$ learn a set of rules to represent mapping f.
- For label L_j on y with prototypes P_j learn a rule IF$(\vec{x}$ is $L_i)$ THEN $(y$ is $L_j)$ where P_i is determined from the set of points $\{\vec{x} : (\vec{x}, y) \in DB$ and $y \in P_j\}$.
- For input $\vec{x}$ then $\mu_{L_j}(y) = \mu_{L_i}(\vec{x}) = \delta([d(\vec{x}, P_i), \infty))$.
- Determine $BetP_y(L_j) = \sum_{F:L_j \in F} \frac{m_y(F)}{|F|}$
- Evaluate $\hat{y} = \sum_{L_j} BetP_y(L_j) c_j$ where c_j is the average of P_j.

Sunspot Database

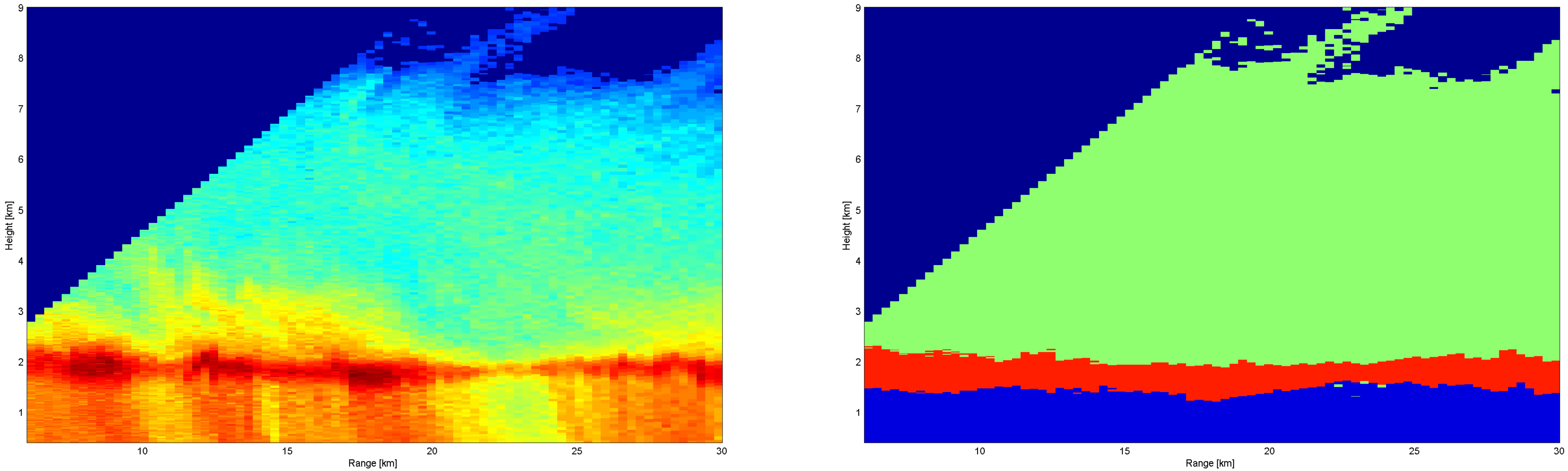


Decision Tree Models



Classification of Weather-Radar Images

- The use of weather radars to estimate precipitation is an increasingly important area in hydrology.
- Brightband= region of enhanced reflexivity cause by amplified echoes due to scattering of microwaves from melting snow



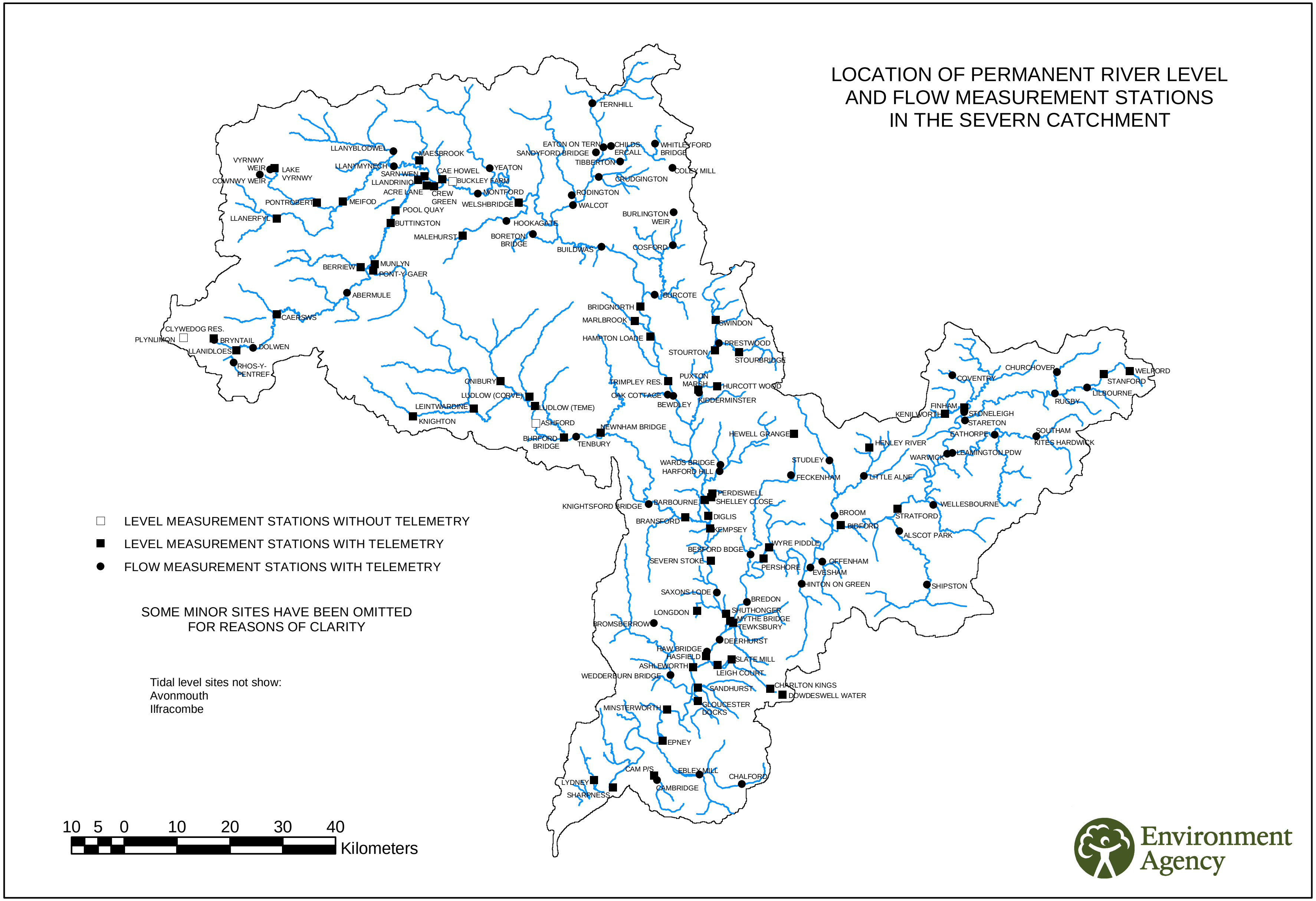
Rules for Radar Classification

- **Attributes:** $Z_h =$ Reflectivity factor, $Z_{dr} =$ Differential reflectivity, $L_{dr} =$ Linear Depolarisation Ration, $H =$ Height of the measurement.
- **Rules:** $(\mathcal{D}_{L_{dr}} = \{low, medium\}) \wedge (\mathcal{D}_H = \{med., high\}) \wedge (\mathcal{D}_{Z_h} = \{high\}) \wedge (\mathcal{D}_{Z_{dr}} = \{med., high\}) \rightarrow Rain : 0.998,\ Snow : 0.002, Brightband : 0$
- $(\mathcal{D}_{L_{dr}} = \{medium\}) \wedge (\mathcal{D}_H = \{med., high\}) \wedge (\mathcal{D}_{Z_h} = \{med., high\}) \wedge (\mathcal{D}_{Z_{dr}} = \{low\}) \rightarrow Rain : 0.03,\ Snow : 0.97, Brightband : 0$
- $(\mathcal{D}_{L_{dr}} = \{high\}) \wedge (\mathcal{D}_H = \{med.\}) \wedge (\mathcal{D}_{Z_h} = \{high\}) \wedge (\mathcal{D}_{Z_{dr}} = \{high\}) \rightarrow Rain : 0.02,\ Snow : 0, Brightband : 0.98$

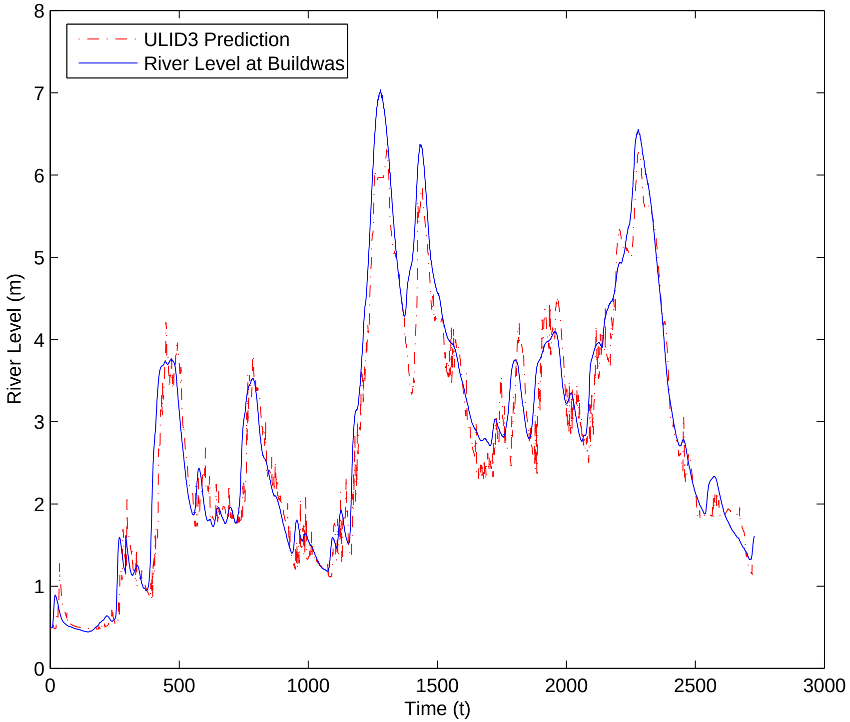
River Severn Forecasting

- The River Severn is situated in the South West of England with a catchment area that spans from the Cambrian Mountains in Mid Wales to the Bristol Channel in England.
- We focus on the Upper Severn from Abermule in Powys and its tributaries, down to Buildwas in Shropshire.
- The Data consists of 13120 training examples from 1/1/1998 to 2/7/1999 and 2760 test examples from 8/9/2000 to 1/1/2001, recorded hourly.
- Each example has 19 continuous attributes falling into two categories; station (water) level measurements and rain fall gauge measurements.
- The forecasting problem is to predict the river level at Buildwas at time $t + \delta$ ($\delta =$ lead time).

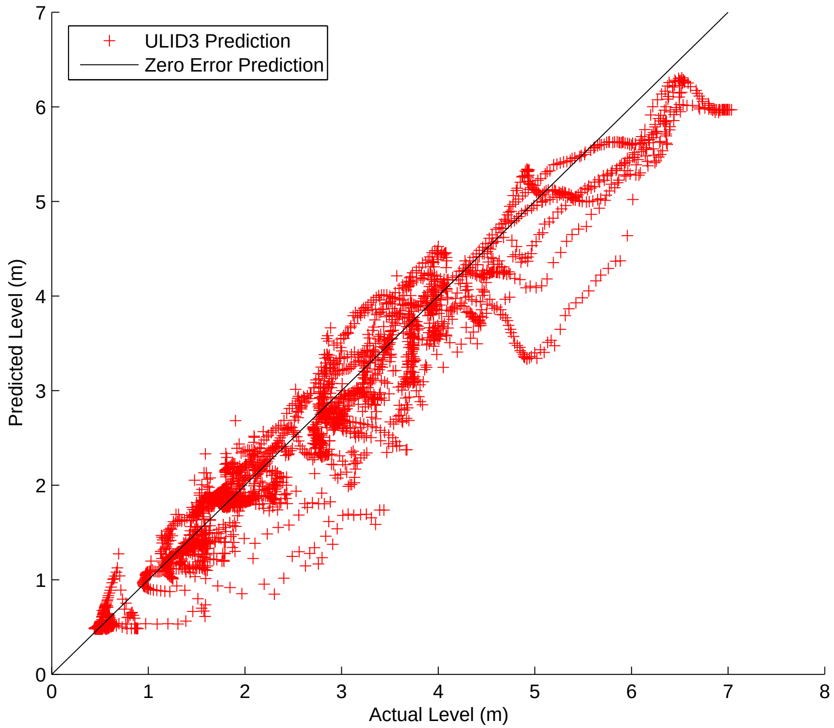
Severn Catchment Map



36 Hours Ahead Prediction



Test Data Set



Scatter Plot

Severn Rules

- x_t^A = time t level at Abermule, x_t^B = time t level at Buildwas, x_t^M = time t level at Meifod
- *Focal set* $\{low\}$: $low \in \mathcal{D}_{x_t^A} \wedge low \in \mathcal{D}_{x_t^B} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{low\}$
- *Focal set* $\{low, medium\}$:
  $low \in \mathcal{D}_{x_t^A} \wedge medium \in \mathcal{D}_{x_t^B} \wedge low \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{low, medium\}$
  $low \in \mathcal{D}_{x_t^A} \wedge medium \in \mathcal{D}_{x_t^B} \wedge medium \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{low, medium\}$
  $medium \in \mathcal{D}_{x_t^A} \wedge low \in \mathcal{D}_{x_t^B} \wedge medium \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{low, medium\}$
- *Focal set* $\{medium\}$:
  $medium \in \mathcal{D}_{x_t^A} \wedge medium \in \mathcal{D}_{x_t^B} \wedge medium \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{medium\}$
- *Focal set* $\{medium, high\}$: $\{medium\}$:
  $high \in \mathcal{D}_{x_t^A} \wedge medium \in \mathcal{D}_{x_t^B} \wedge medium \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{medium, high\}$
  $high \in \mathcal{D}_{x_t^A} \wedge medium \in \mathcal{D}_{x_t^B} \wedge high \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{medium, high\}$
  $high \in \mathcal{D}_{x_t^A} \wedge high \in \mathcal{D}_{x_t^B} \wedge medium \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{medium, high\}$
  $medium \in \mathcal{D}_{x_t^A} \wedge high \in \mathcal{D}_{x_t^B} \wedge high \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{medium, high\}$
- *Focal set* $\{high\}$: $high \in \mathcal{D}_{x_t^A} \wedge high \in \mathcal{D}_{x_t^B} \wedge high \in \mathcal{D}_{x_t^M} \rightarrow \mathcal{D}_{x_{t+36}^B} = \{high\}$

Conclusions

- Granular modelling aims to provide high-level linguistic (rule-based) models for complex systems.
- The approach balances the requirements of predictive accuracy and model transparency.
- We have proposed random set approach to model imprecise labels consistent with an epistemic theory of vagueness.
- A prototype theory interpretation has been also introduced.
- A number of case study applications of granular modelling have been described.

Shameless Advertising

- *Modelling and Reasoning with Vague Concepts* by Jonathan Lawry, Springer 2006