ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Classification of non-meteorological targets with new statistical methods

Jarmo Koistinen, Teemu Mäkinen,
and Seppo Pulkkinen
Finnish Meteorological Institute/Earth Observation

Empirical automatic classification

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

1. Manually select and classify a number of cases
2. Derive statistical models to represent trained class data sets
3. Find the optimal set of separating measured or derived quantities for each class
4. Determine the probability of a measurement bin to belong to a given class applying steps 2-3 above
5. Perform validation with independent data
6. Run the classification for real time products

Measured data

Conventional moments (Doppler radar, H)

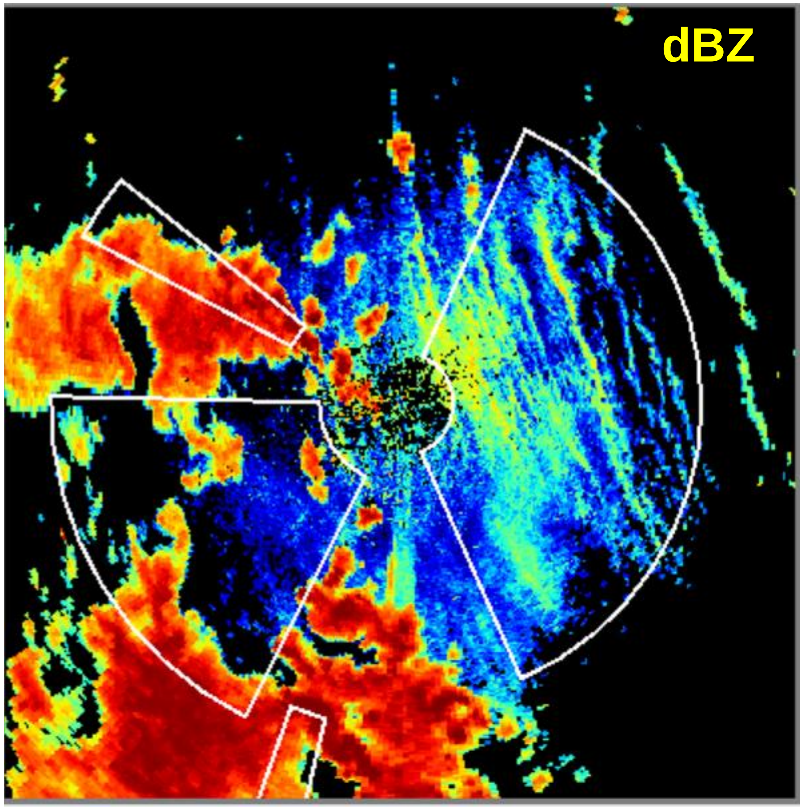
- dBT (no filtering), dBZ (clutter filtering)
- V, W, SQI (phase coherency)

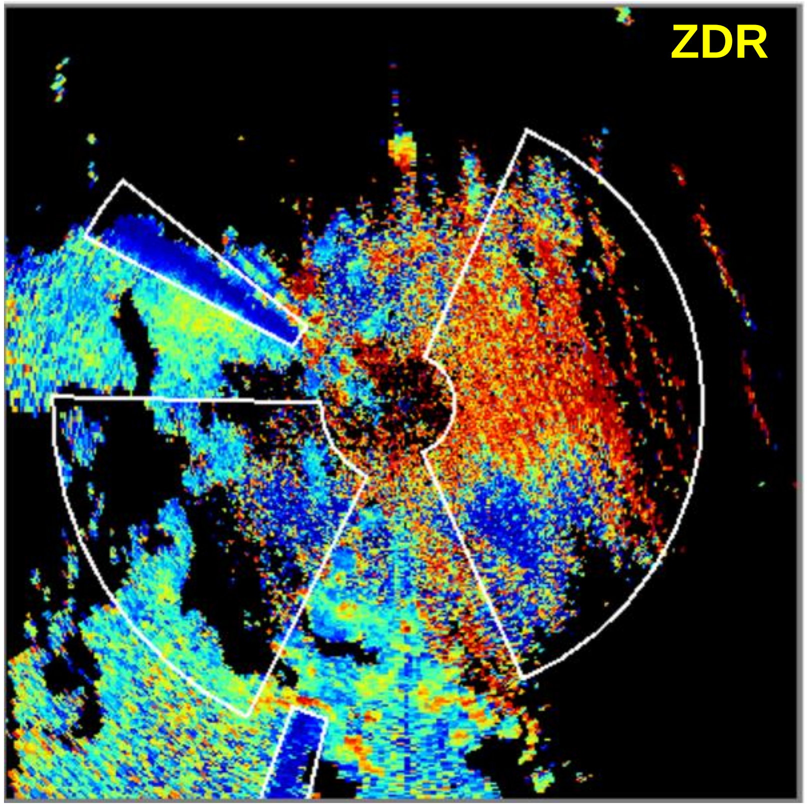
Dual pol moments (transmit H&V, receive H&V)

- dBZE (enhanced reflectivity factor from H & V)
- ZDR, Differential Reflectivity (Factor)
- ρ_{HV} , (Copolar) Correlation Coefficient
- Φ_{DP} , Differential Phase (Shift)
- [K_{DP} , Specific Differential Phase (Shift)]
- DDV, Differential Doppler Velocity, V_H-V_V

Each quantity (9) adds a dimension to a data **sample**

Pure single class segments are selected from PPI images using a graphical user interface (thresholding possible)

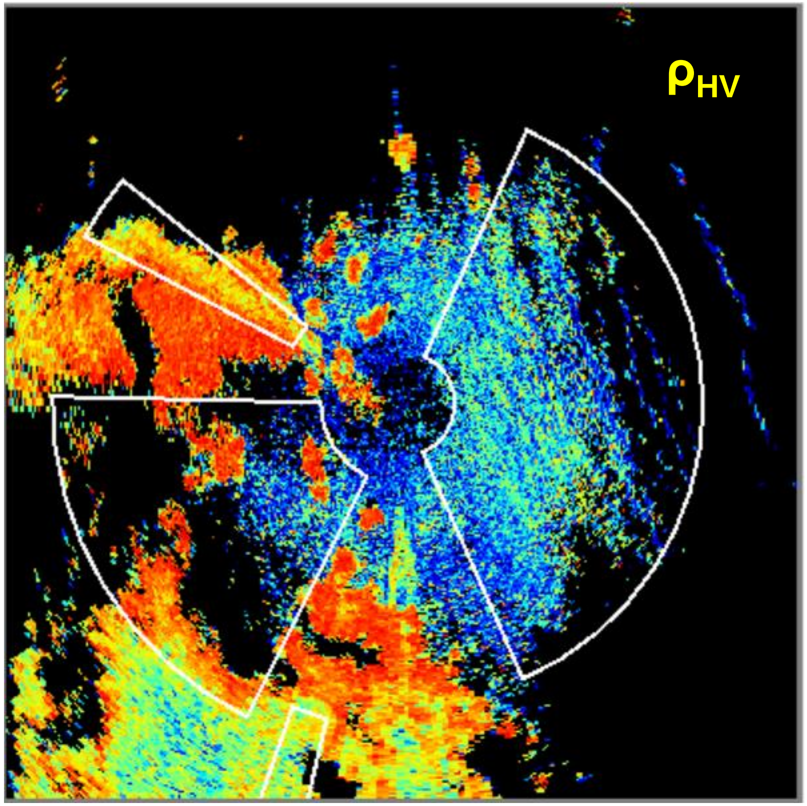


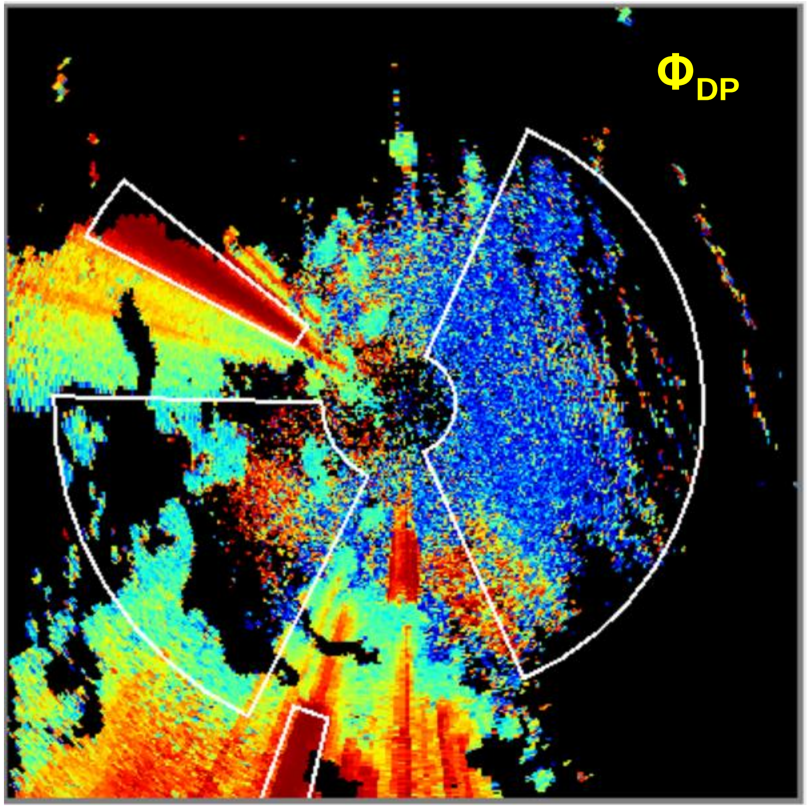


Example ANJ radar 17 May 2012,14:15 UTC. Selected cases: insects_day_land, rain_convective, rain_attenuated.

Success criteria for training

- The teacher(s) should be unbiased and consistent
- The phenomena should be consistent in time and space
- The radars should be unbiased and homogeneous
- The data sample for each class should be representative





Classes applied

In total some 1000 cases used in the training:

- 3-5 types of solid precipitation
- 5-6 types of liquid precipitation
- 2-3 types of melting precipitation
- 4 types of insects
- 4-5 types of birds
- 1 insects blended with birds
- 4 types of ground and sea clutter
- 4 types of anthropogenic echoes & the sun
- 3 types of radar system effects

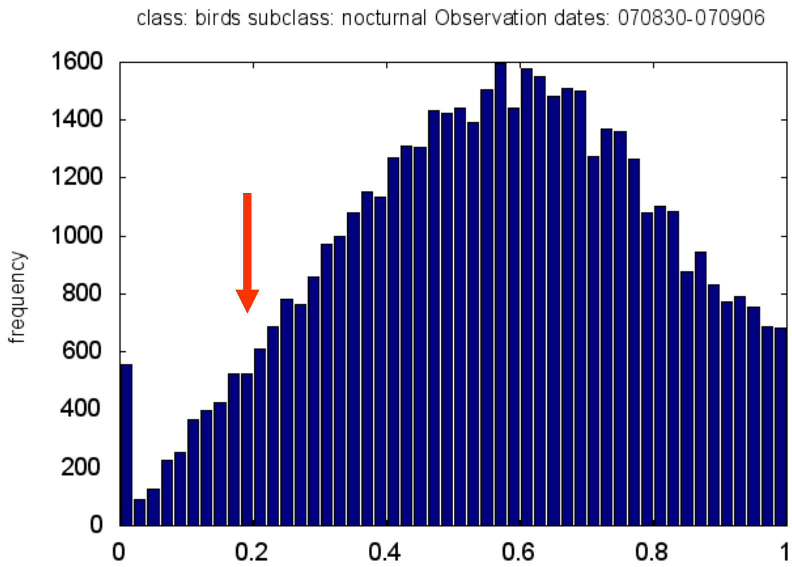
Overlapping of fine-grained classes does not matter if the model allows formation of union classes without retraining.

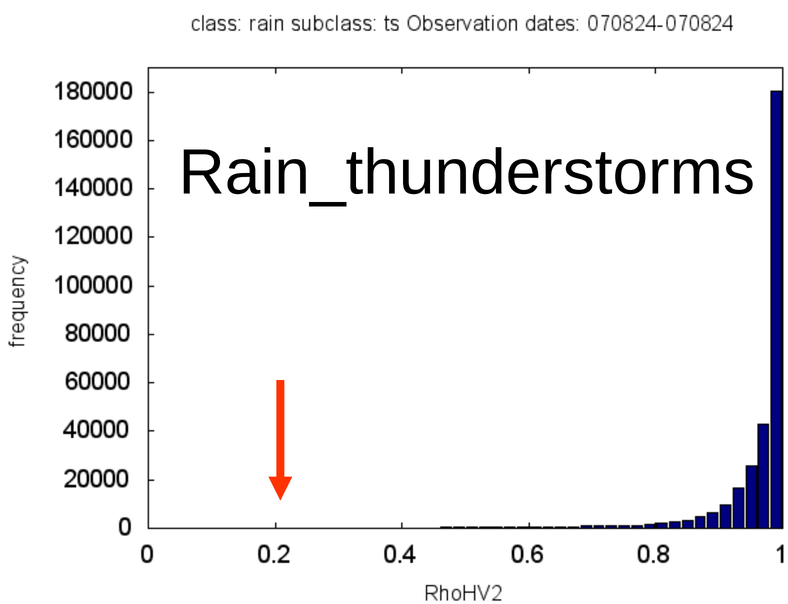
ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

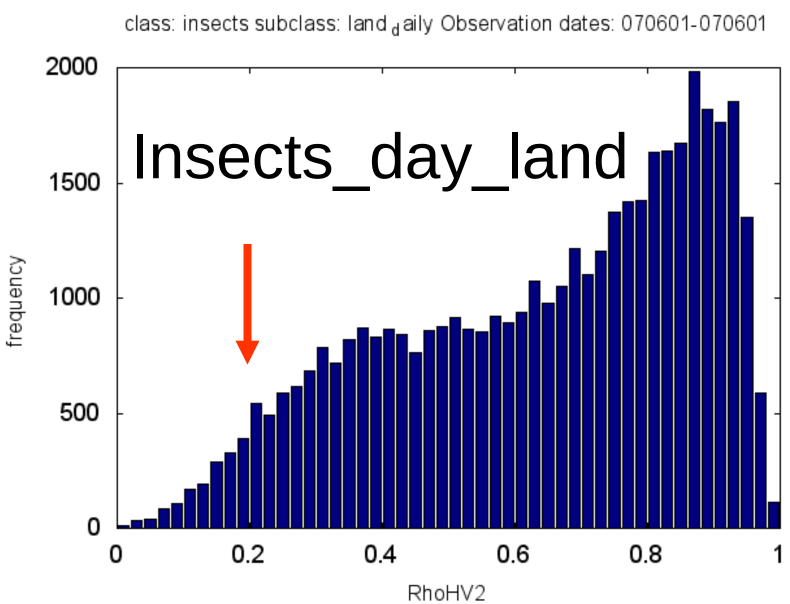
Separation with PDFs

Available classifications applying 1D - 3D membership functions of the measured quantities separate quite well meteorological data points from non-meteorological ones. They can't separate each non-meteorological target class. Example: $\rho_{HV} = 0.2$

Birds_nocturnal







Enhancement of separation skill

ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Texture filtering: A human views patterns of the measured quantities, not so much point values.

- Provides additional dimensions for optimization
- We use a 9x9 bin area to derive non-local information
- Filter set selected to minimize radar specific properties like absolute values of measured moments
- Currently a set of 7 common filters have been used:
  - **Availability $\mathcal{A}$ (dBT) & Border $\mathcal{B}$ (dBT)**
  - **Anisotropy $\mathcal{S}$ (*), Gamma $\mathcal{G}$ (*), Order/Entropy $\mathcal{O}$ (*), Granularity $\mathcal{R}$ (*), Variance $\mathcal{V}$ (*) (*** = all moments**)**

Challenging implication for statistical modeling: Each data sample has approximately 50 dimensions, which exhibit often non-Gaussian PDFs

Method 1*: Principal component analysis (PCA)

- The model can be optimized to
  - **Maximize the average separation skill of all classes**: **a general model** (usually large number of dimensions)
  - **Maximize the detection skill of a wanted class: a dedicated model** (usually 5-10 dimensions is enough)
- PDF of data is described by a Gaussian volume called hyperellipsoid obtained through the PCA algorithm
- Proper statistical treatment of incomplete data and cyclic quantities (non-standard PCA) implemented
- Provides well defined metrics for class membership and class intersections and unions
- Still a fast and simple method
- **Limitation: Assumes Gaussian PDFs**

* Used in 2009-2011 (see ERAD 2010 presentations), training data 350 cases

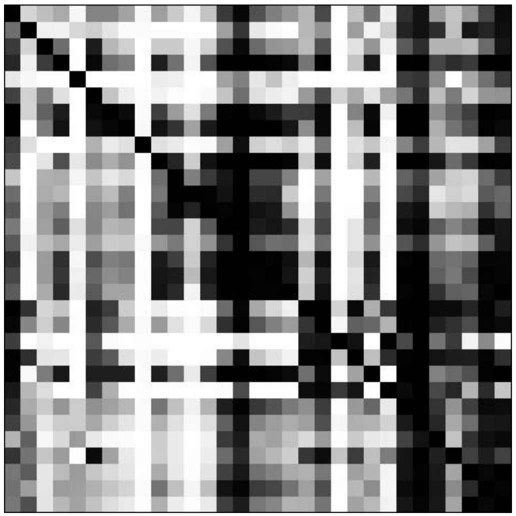
Evaluation of general models with intersection matrices

Black denotes cross talk between classes, white separation

Doppler Radar without texture

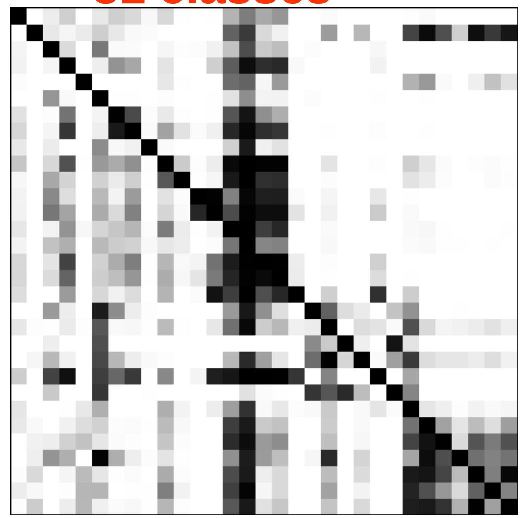
31 classes

31 classes



31 classes

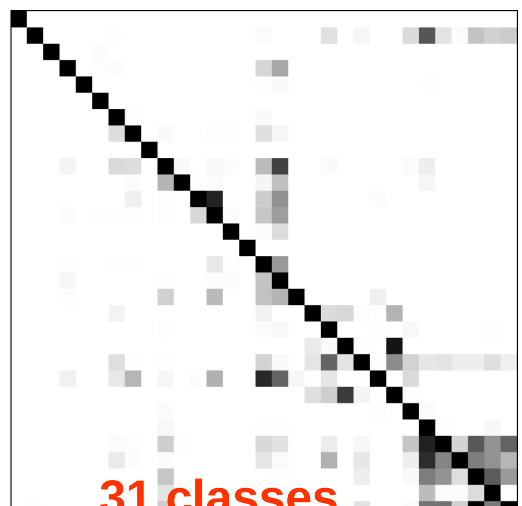
31 classes



Dual pol radar without texture full model

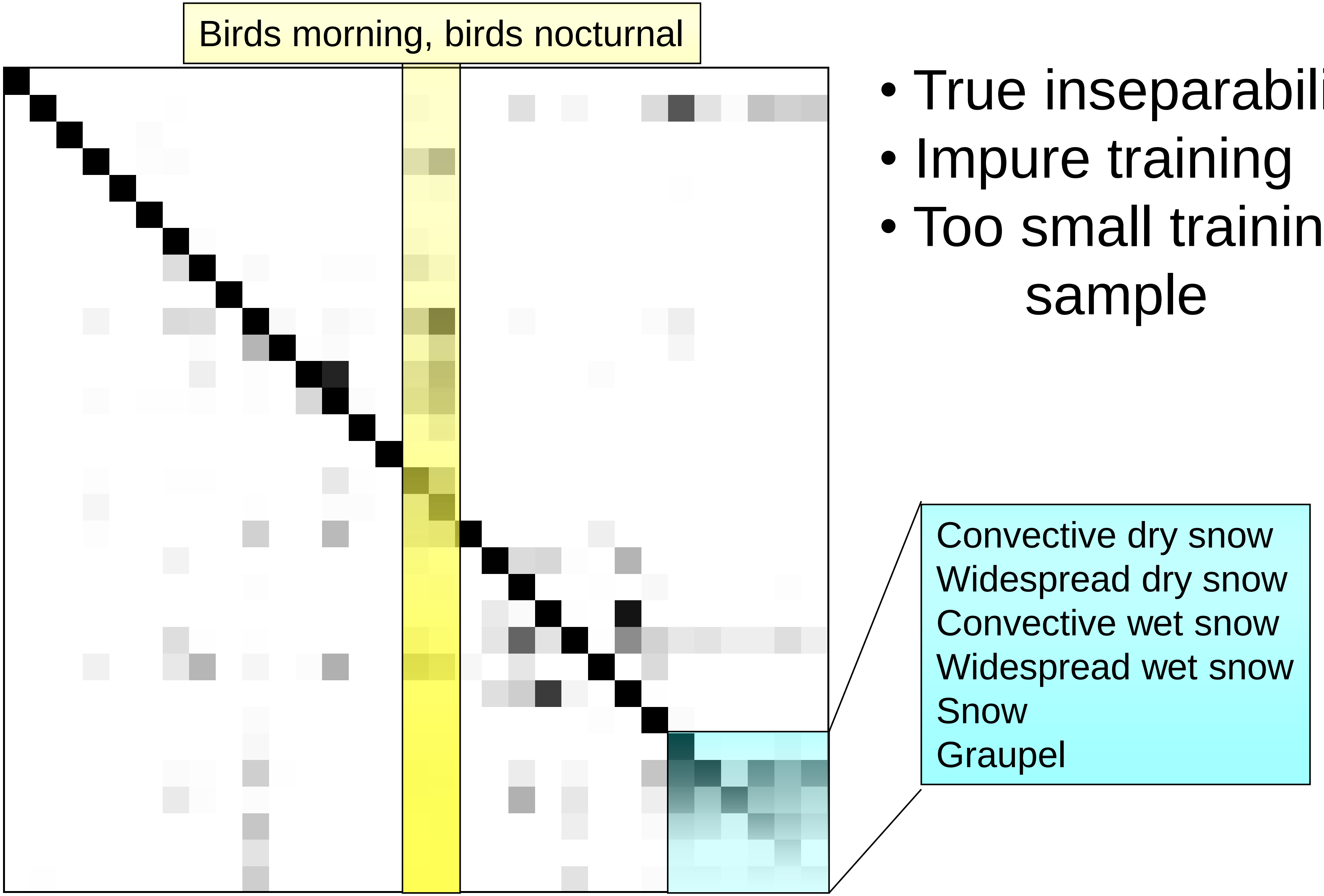
Dual pol radar with texture, the best 10 D model (↑)

31 classes



Dual pol radar with texture, full model

Reasons of less good separation skill



- True inseparability
- Impure training
- Too small training sample

Method 2*: Orthogonal polynomial transfer (OPT)

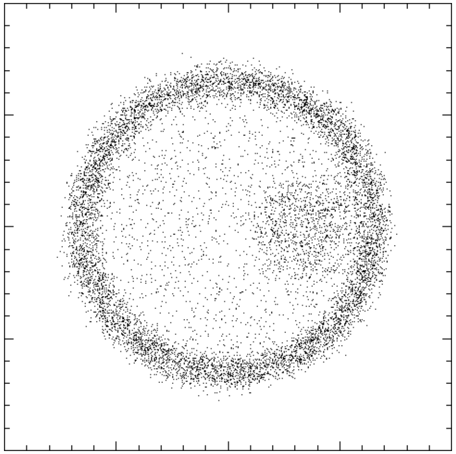
- New method, developed in this effort. Finds the subspace with minimum overlapping of class PDFs
- Based on orthogonal polynomials in normalized coordinates
- Computationally fast (no iteration), models are parallelizable => no retraining needed for class unions
- Models directly the PDFs of (almost) any shape
- Projection to separable sub-spaces trivial ⇔ Easy to find the best separating dimensions

* Mäkinen and Holmström: Modeling Probability Density through Ultraspherical Polynomial Transformations (to be submitted).

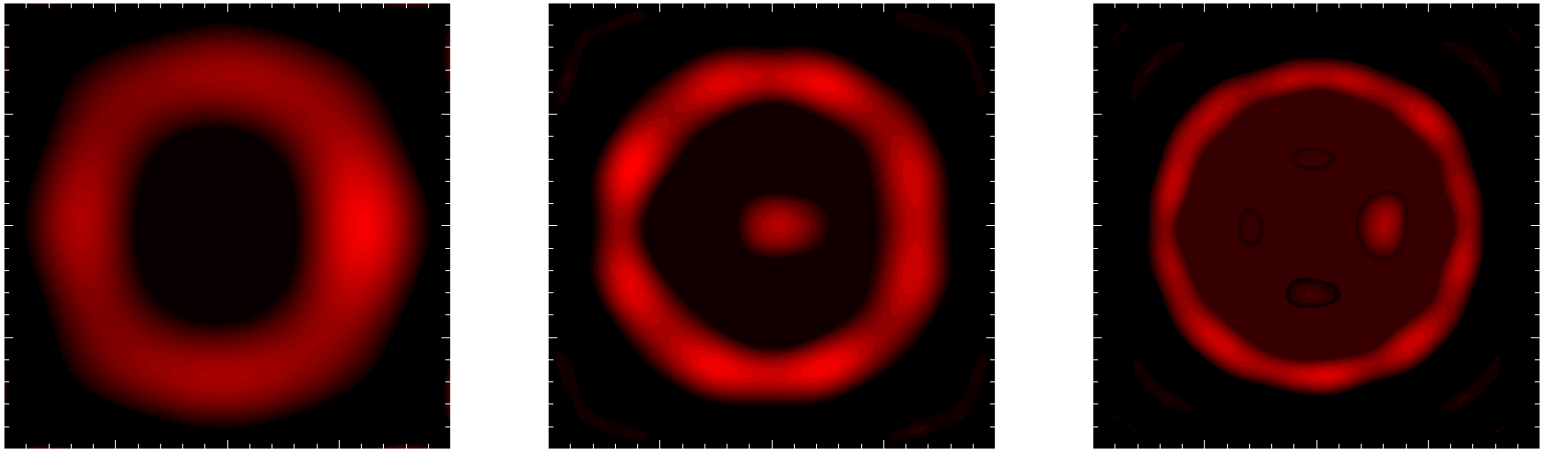
Using Legendre polynomials for PDF

$$p(x) \approx \sum_n c_n P_n(x)$$

$$c_n = \frac{2}{2n+1} \int dx P_n(x)\, p(x) \approx \frac{2}{2n+1} \sum_i P_n(x_i)$$

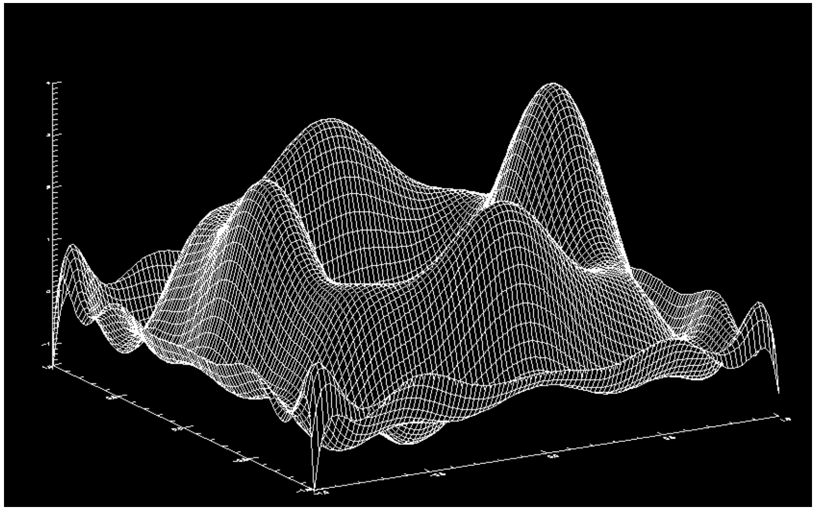


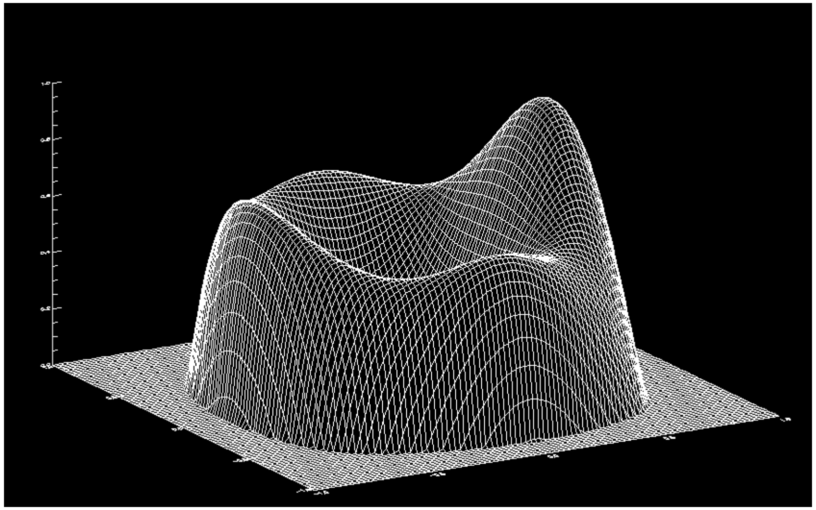
n = 8 n = 15 n = 25



Requires an analytic model fit to the expansion

- Any function can be easily turned into a *Chebyshev* series through a Discrete Cosine Transformation (DCT)
- A Chebyshev series can be converted to a Legendre series through a L-space transformation
- Two Legendre series can be matched through χ^2 fitting
- From the algorithm POV the model is a black box – **nothing is assumed about the actual form of the PDF**





Normalization of a class probability

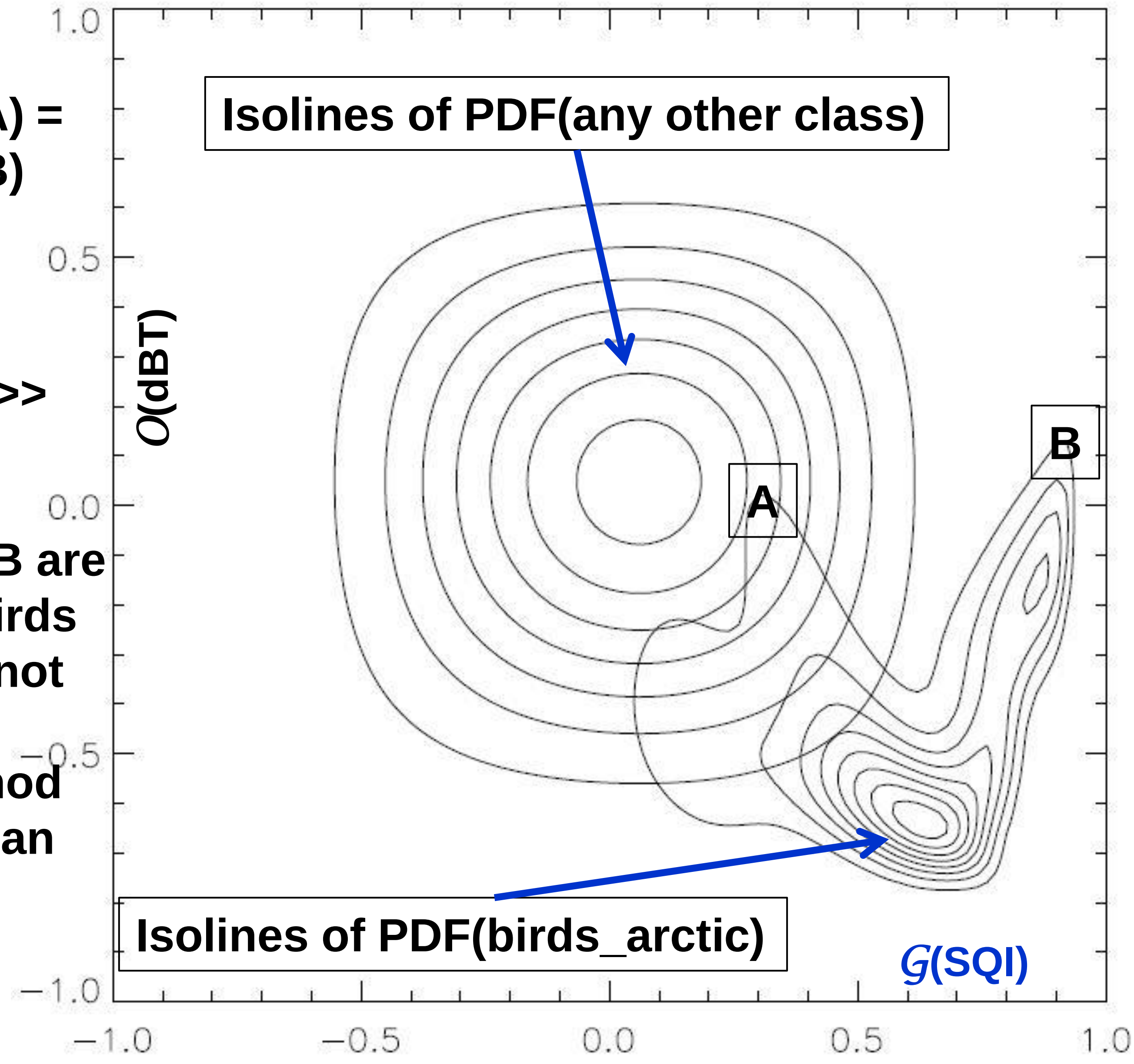
PDF(birds_arctic,A) = PDF(birds_arctic,B)

but

PDF(any_other,A) >> PDF(any_other,B)

hence samples at B are very likely arctic birds but those at A are not

Note how the method allows non-Gaussian class PDFs



Class separators

Class	Cases	Bins	Separating sub-space
aircraft	62	709	SQI $\mathcal{V}$(W)
birds_arctic	89	181 869	SQI Φ_{DP} $\mathcal{O}$(dBZ)
birds_day	12	22 034	SQI $\mathcal{G}$(SQI) $\mathcal{V}(\Phi_{DP})$
birds_insects	19	344 412	SQI V $\mathcal{O}$(W)
birds_morning	47	290 359	SQI $\mathcal{V}$(V) $\mathcal{O}$(dBZ)
birds_nocturnal	84	1 306 795	SQI V $\mathcal{O}$(W)
buildings	25	365 433	SQI V $\mathcal{R}$(dBT)
chaff	14	2 845	V $\mathcal{S}$(SQI) $\mathcal{V}(\Phi_{DP})$
emitter	71	96 698	Φ_{DP} $\mathcal{V}$(V)
ground_anaprop	63	259 611	SQI $\mathcal{A}$(dBZ) $\mathcal{O}$(dBT)
insects_day_land	49	1 028 681	SQI V $\mathcal{R}$(dBZ)
insects_day_sea	20	153 112	SQI V Φ_{DP}
insects_night_land	14	197 910	SQI V $\mathcal{R}$(dBZ)
insects_night_sea	2	15 445	SQI $\mathcal{V}(\Phi_{DP})$
noise	37	4 453	$\mathcal{V}$(dBT) $\mathcal{V}$(V)
rain_Cb	46	481 202	dBT Φ_{DP} $\mathcal{R}$(dBZ)
rain_Ns	24	365 750	dBT Φ_{DP} $\mathcal{R}$(dBZ)
rain_Ts	4	10 736	dBT Φ_{DP} ρ_{HV}
rain_attenuated	25	74 560	dBT $\mathcal{O}(\Phi_{DP})$ $\mathcal{S}$(dBT)
reflection	6	1 188	$\mathcal{S}$(dBT) $\mathcal{S}$(V)
sea_sausage	23	57 869	SQI V $\mathcal{V}(\Phi_{DP})$
sea_sector	13	26 264	SQI $\mathcal{G}(\Phi_{DP})$ $\mathcal{O}(K_{DP})$
second_preci	28	30 709	SQI V $\mathcal{O}(\Phi_{DP})$
ships	28	16 581	SQI Φ_{DP} $\mathcal{V}(\Phi_{DP})$
sidelobe	4	229	$\mathcal{B}$(dBZ) $\mathcal{O}$(dBT) $\mathcal{S}$(dBT)
sleet_Cb	30	199 426	dBZ Φ_{DP} $\mathcal{G}$(SQI)
sleet_Ns	38	469 766	dBT Φ_{DP} $\mathcal{R}$(dBZ)
snow_Cb	32	305 272	$\mathcal{R}$(dBZ) $\mathcal{S}(K_{DP})$ $\mathcal{G}$(SQI)
snow_Ns	53	1 379 795	dBT SQI $\mathcal{G}(\rho_{HV})$
sun	37	20 650	$\mathcal{S}$(dBT) $\mathcal{V}$(V)

The most frequent significant separators

<u>All classes</u>:

SQI

Φ_{DP}

V

dBT

$\mathcal{R}$(dBZ) or $\mathcal{R}$(dBT)

<u>Non-meteorological</u>

SQI

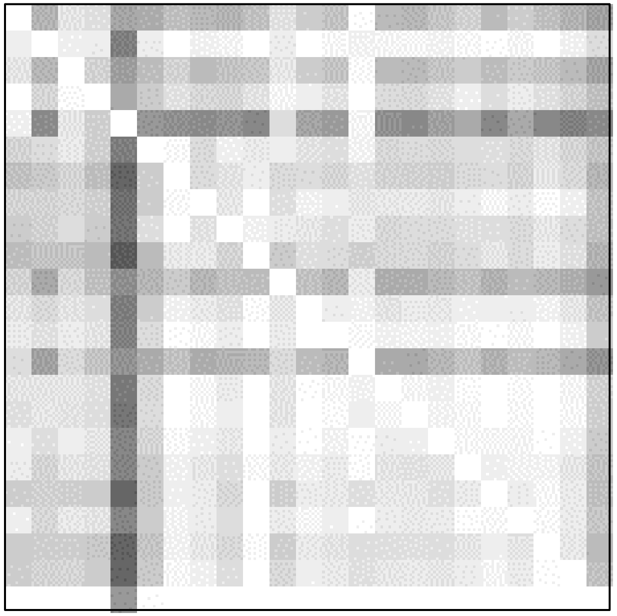
V

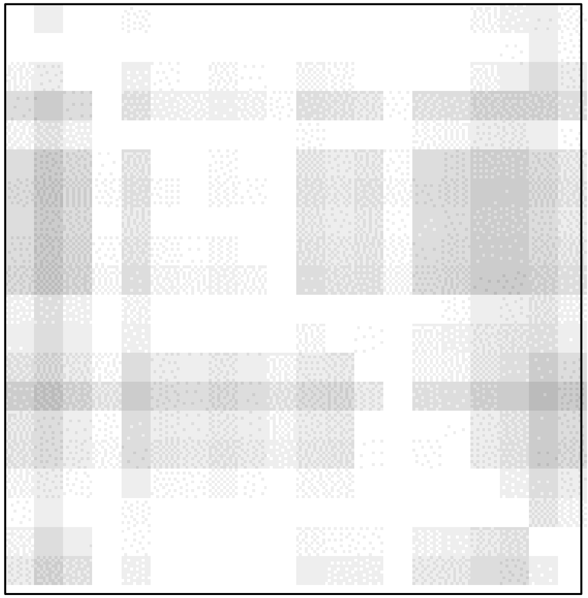
Φ_{DP} or $\mathcal{V}(\Phi_{DP})$

$\mathcal{R}$(dBZ)

Testing of case homogenity

Homogenity of the training cases can be tested using the confusion matrix of all cases in each class. PDFs not matching well to the class-averaged PDF-model easily detected for further evaluation.

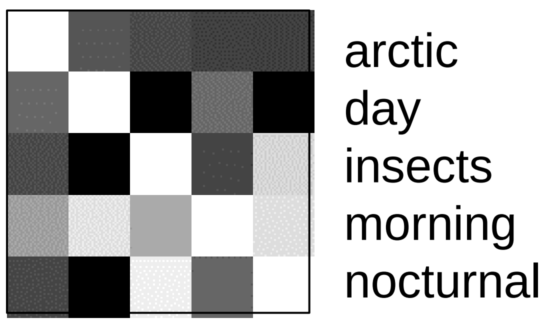




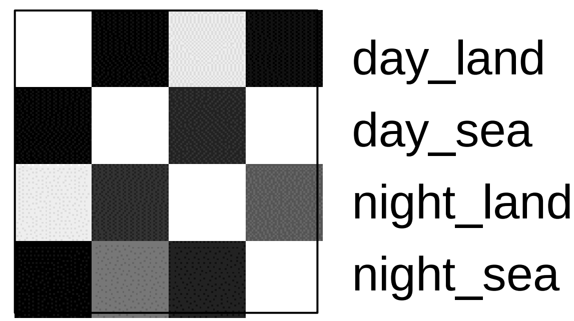
Dedicated models for sub-classes

- Dedicated optimized models can be generated for classes (sub-classes) which are near to each other. In such cases the model contains two phases.
- The method and reasonability of fine-grained classes must be considered case by case.

- Examples below: Original bird classes are moderately separable but insect classes are not.



Bird classes



Insect classes

Summary

- 1000 ”pure” PPI segments were classified manually to belong to ~30 meteorological and non-meteorological target classes.
- Principal component analysis (PCA) can be used but it has limitations solved by a new method called orthogonal polynomial transfer (OPT)
- Reflectivity, Doppler and dual pol quantities as well as texture properties of them are all crucial components among the best separating variables.
- The class separation skill, after integration of some too similar classes, is very good in the dependent data but not yet tested with independent cases.
- Multidimensional (~50) PCA and OPT are slow to train but fast in action.
- Automatic target classification provides useful information for improved QPE QC and for many applications.