

# Computing With Perceptions for the linguistic description of complex processes through time series data analysis

Generating linguistic descriptions of data



Alejandro Ramos Soto

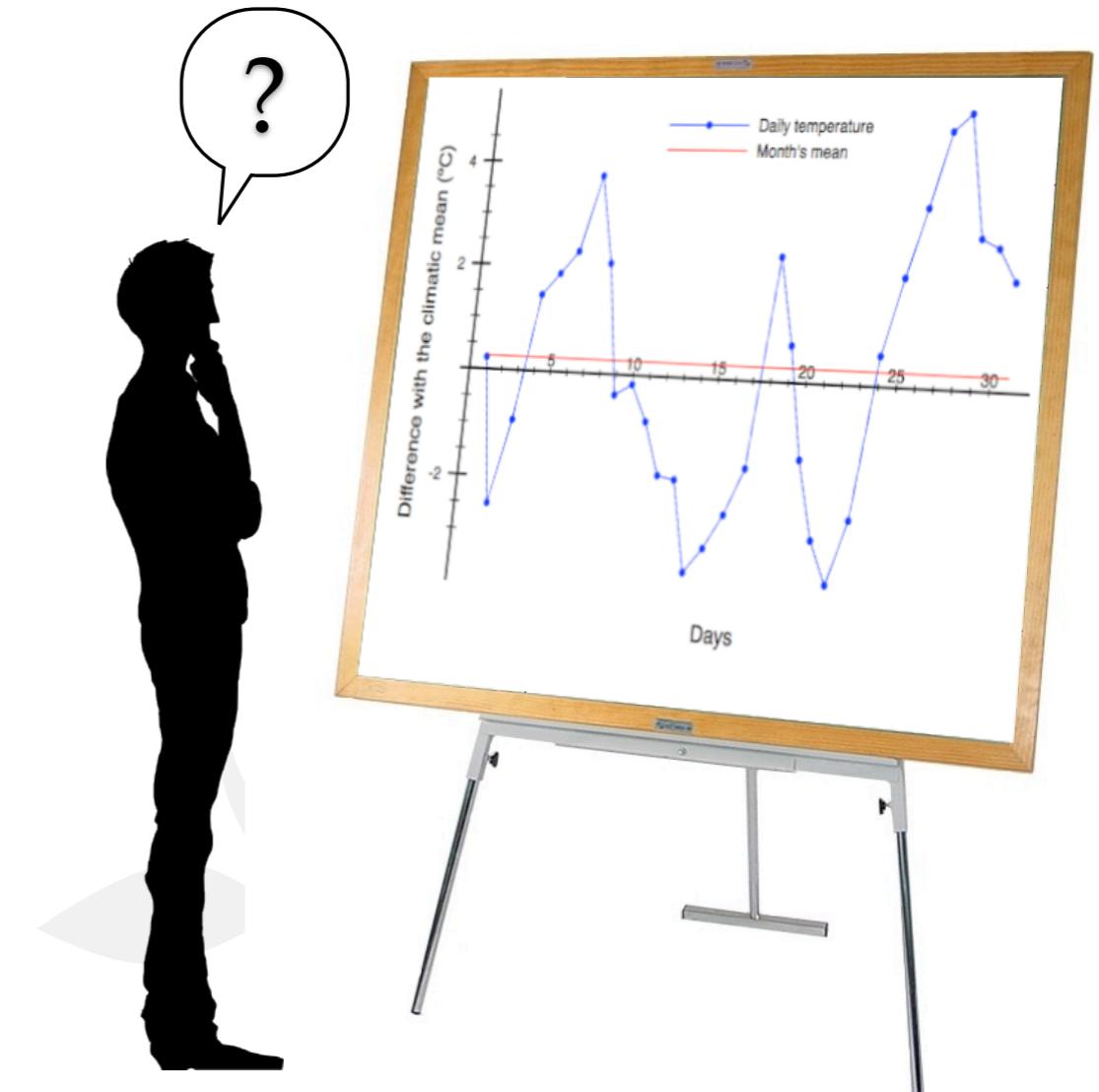
PhD Supervisors: Alberto Bugarín, Senén Barro

[citus.usc.es](http://citus.usc.es)

# The need for information

## The era of data

- Nowadays huge quantities of data are continuously generated (sensor architectures, web, financial markets...)
- Big Data has recently acquired great relevance
- People can't assimilate raw data
- We need predictive, exploratory or descriptive techniques to understand data



# The need for information

## Information gathering techniques

Several fields explore this task:

- Traditional statistic techniques
- Data mining:
  - Classification
  - Association
  - Clustering
  - Trend analysis
- However, in many cases the extracted information is not adapted to its final users: people

# The need for information

## Information gathering techniques

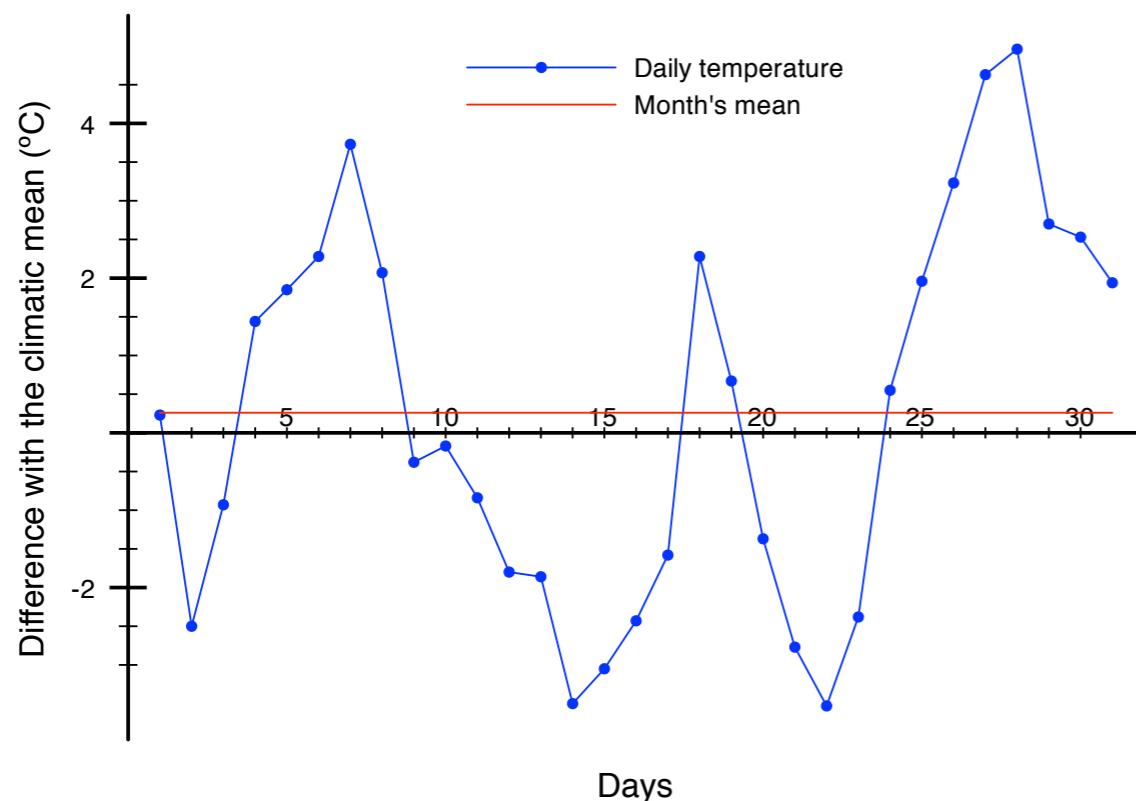
Other techniques follow a more human-friendly approach, providing information through linguistic concepts and natural language texts:

- Soft-computing field:
  - Linguistic descriptions of data:
    - Provide a general, short and precise description of numeric data series, using linguistic labels
    - Use fuzzy sets to model the imprecision inherent to linguistic terms
  - Natural language generation (NLG) field:
    - Set of methodologies and techniques which generate texts from data

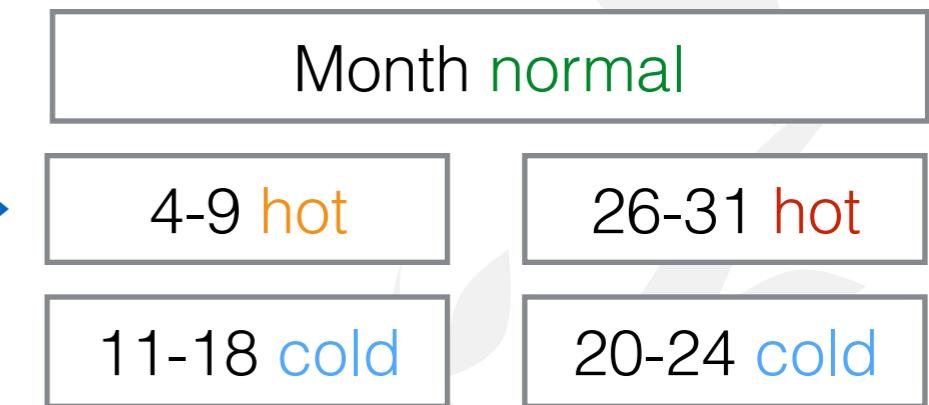
# The need for information

Linguistic descriptions of data and natural language generation

Linguistic descriptions of data and natural language generation are complementary fields:



linguistic  
description



natural language  
generation

***The month was normal in average, with low temperatures during the second and third weeks and very high values in the last week.***

# Linguistic descriptions of data

## Application fields

There are many application domains in which these kind of techniques may prove useful for both expert and non-expert users. Some examples:

- Meteorology (our current research domain)
- Economy: “*Towards the end of the session the prices dropped*”
- Domestic consumption: “*Most of the days the midday electric consumption was higher than night consumption*”
- Project management reporting: “*The project is in a critical state, since the funds are running out and we expect to complete the current tasks in six months*”
- Industrial monitoring: “*The valves have been operating correctly for the previous two hours, although a high pressure event was detected half an hour ago*”

# PhD Hypothesis

Linguistic description techniques combined with natural language generation approaches have the potential to generate high quality textual descriptions from complex data and thus to fulfill human information needs

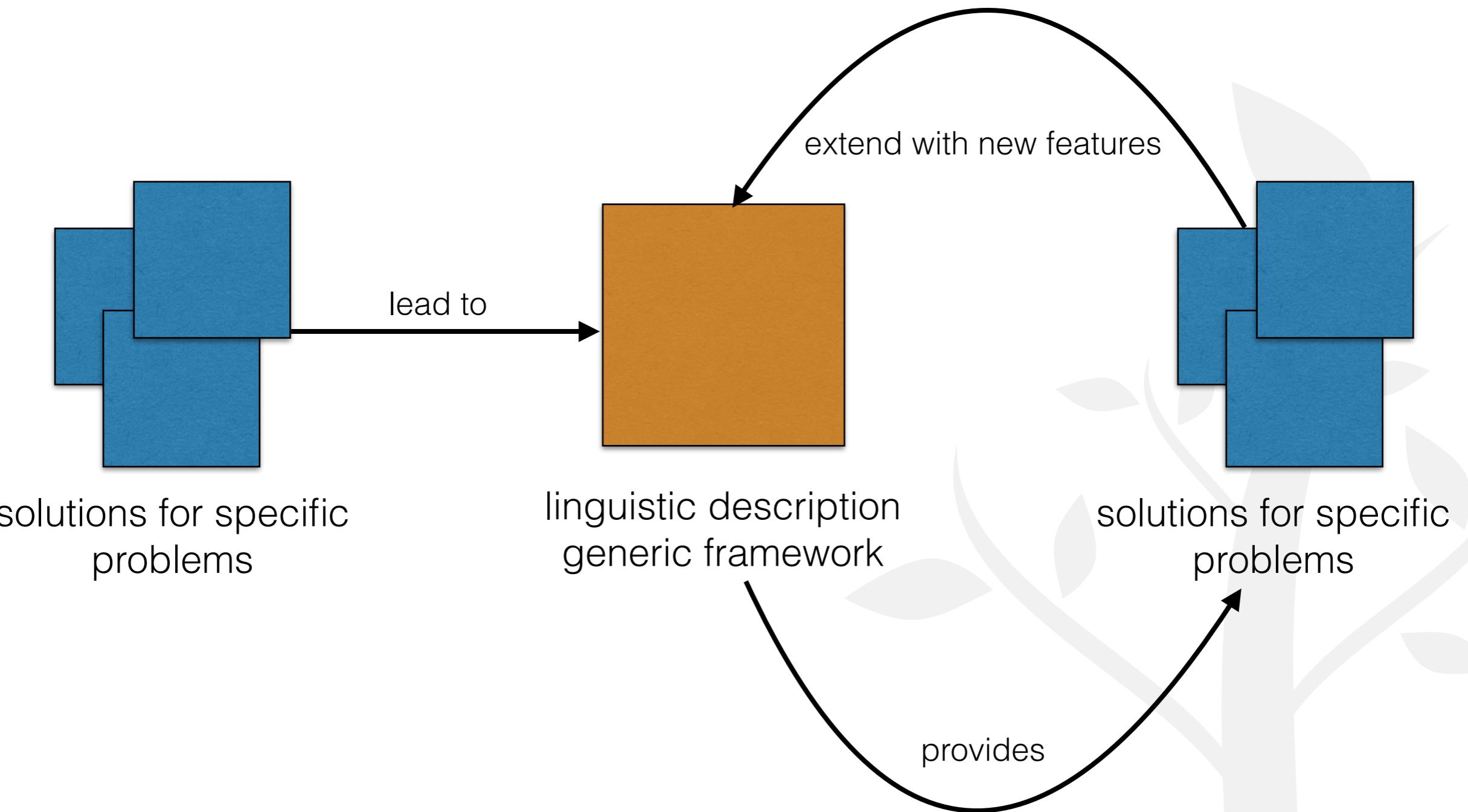
# PhD Tasks and Objectives

## What

- To **research** on **linguistic description generation techniques**, including:
  - Fuzzy evaluated and quantified sentences
  - Trends
  - Computing in **time** and **space** dimensions
- To **solve** linguistic description **problems** to **satisfy** real life **information needs**:
  - To identify application fields
  - To apply linguistic description techniques in these domains
  - To develop natural language approaches for these solutions
- Ultimately, to **design** and **develop** a **generic** linguistic description generation **framework**

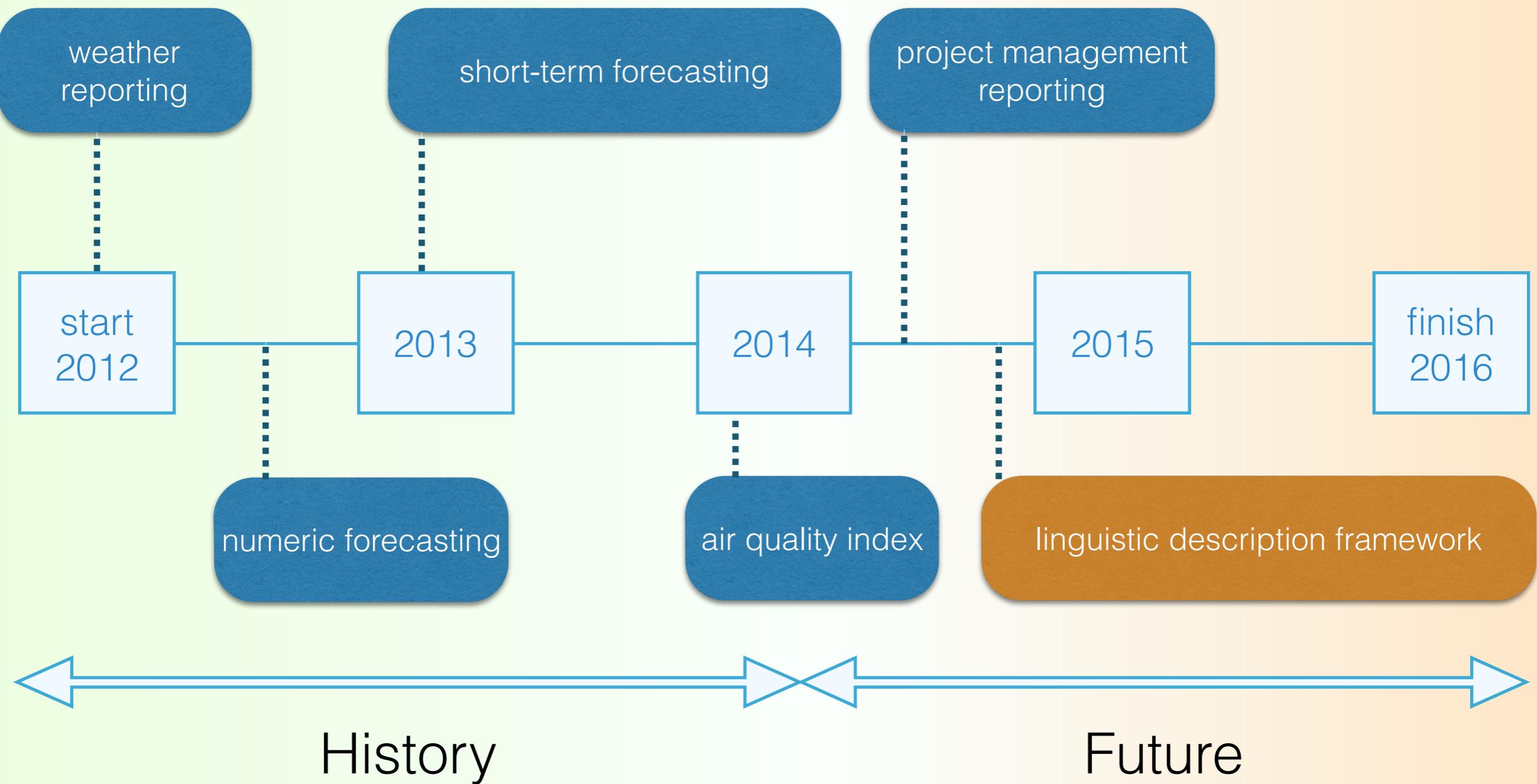
# PhD research methodology

How



# PhD Chronology

When



# Linguistic Descriptions in Meteorology

Meteorology: A great domain for experimentation...

Meteorology is a rich domain for generating linguistic descriptions from data:

- Huge quantities of real observation data from meteorologic stations
- Huge quantities of prediction data from numeric simulation models
- ...in both time and space dimensions

Spain - Los Lances (wave: NWW3 10.9. 2013 00 UTC)		[Options]																															
Forecast	2D	Map	Webcams	Wind reports	Accommodation	Schools/Rentals	Shops	Other...																									
GFS	Tu	Tu	Tu	Tu	Tu	We	We	We	We	We	Th	Th	Th	Th	Fr																		
10.09.2013	10.	10.	10.	10.	10.	11.	11.	11.	11.	11.	12.	12.	12.	12.	13.	13.	13.	13.	13.	13.	13.	13.	13.	13.	13.	13.							
00 UTC	05h	08h	11h	14h	17h	20h	05h	08h	11h	14h	17h	20h	05h	08h	11h	14h	17h	20h	05h	08h	11h	14h	17h	20h	05h	08h	11h	14h	17h				
Wind speed (knots)	6	6	8	10	10	8	7	7	9	12	12	11	13	12	15	17	17	13	10	9	12	14	13	13	13	13	13	13	13				
Wind gusts (knots)	8	9	12	13	14	12	12	11	14	17	18	20	24	23	22	23	23	23	20	19	19	19	19	19	19	19	19	19	19				
Wind direction	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←	←				
Wave (m)	0.6	0.7	0.7	0.7	0.7	0.6	0.7	0.7	0.7	0.7	0.7	0.8	1.3	1.5	1.5	1.3	1.2	1.3	1.5	1.5	1.3	1.2	1	1	1	1	1	1	1	1			
Wave period (s)	7	7	8	8	8	7	7	7	7	7	7	8	9	9	9	9	6	6	6	6	6	6	6	6	6	6	6	6	6	6			
Wave direction	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓			
*Temperature (°C)	20	20	24	26	26	23	20	20	24	26	25	23	22	22	24	25	25	23	22	21	24	25	25	25	25	25	25	25	25	25			
Cloud cover (%) high / mid / low													69	56	31	21					8	11	48										
*Precip. (mm/3h)													30	29	39	38	29	16	30	22	21	22											
Windguru rating	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★	★

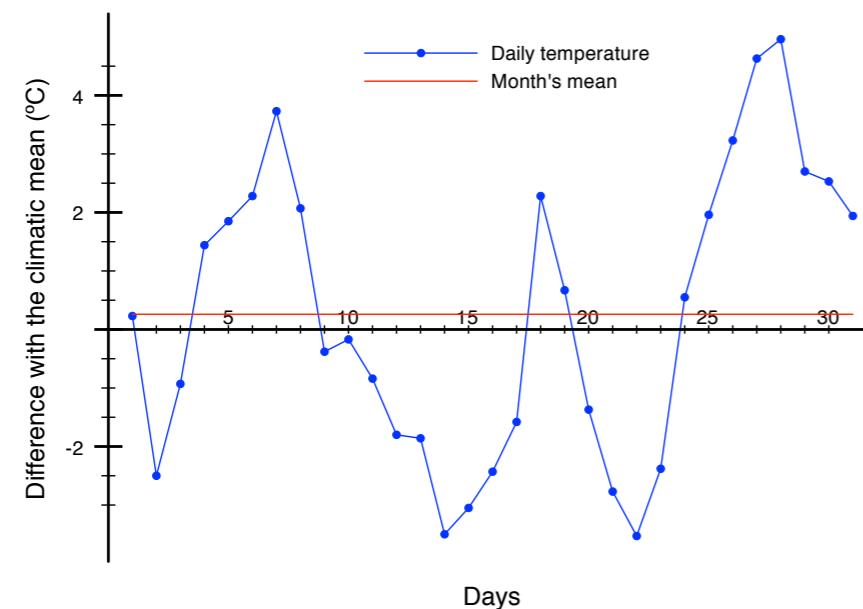


# Linguistic Descriptions in Meteorology

...and for real linguistic description needs and use cases

## CLIMATE REPORTS

- Provide monthly descriptions of the climatic behavior
- *“The temperature was normal for June, with very cold temperatures during the second week and very high during the first and fourth week”*



## NUMERIC FORECAST DESCRIPTIONS

- Linguistic descriptions from numeric prediction data (text prediction)
- *“In the morning the sky will be cloudy or covered, whereas in the evening the amount of clouds will decrease and it will be just cloudy”*



# Linguistic Descriptions in Meteorology

...and for real linguistic description needs and use cases

## SHORT-TERM WEATHER FORECASTS

- We have developed a linguistic description solution for short-term weather forecasts
- A more specific case of numeric prediction, which covers the 315 Galician municipalities
- It describes the weather behavior for the nearest three or four days

xoves, 18 de xullo			venres, 19 de xullo			sábado, 20 de xullo			domingo, 21 de xullo		
Mañá	Tarde	Noite	Mañá	Tarde	Noite	Mañá	Tarde	Noite	Mañá	Tarde	Noite
←	↗	↔	↗	↗	○	○	↗	○	○	↖	↓
5%	5%	5%	5%	15%	5%	5%	40%	20%	35%	5%	5%
MIN 14°	MAX 33°		MIN 16°	MAX 31°		MIN 15°	MAX 29°		MIN 15°	MAX 24°	
18/07/2013 09:00			18/07/2013 10:00			18/07/2013 11:00			18/07/2013 11:00		

*The period will start with clear skies but towards the end it will be rather cloudy. The maximum temperatures will be high but they will decrease moderately towards the end, while the minimum will remain without changes.*

# Linguistic Descriptions in Meteorology

Real use case: Short-term weather forecasts

- MeteoGalicia's website ([www.meteogalicia.es](http://www.meteogalicia.es)):
  - 2 million visits per month
  - 700 thousand visits to the municipalities section per month

### Predición para concellos

**Concellos**

- A Coruña
- Ferrol
- Lugo
- Ourense
- Pontevedra
- Santiago de Compostela
- Vigo

**Buscar outro concello**

Seleccione provincia:

- A Coruña
- Lugo
- Ourense
- Pontevedra

Seleccione concello de A Coruña:

**Os meus concellos**

Listado

Na lista superior dispón de un acceso directo á predición das principais cidades galegas. Pode engadir a esta lista cinco concellos máis seleccionando un novo concello na opción de "Buscar outro concello" e premendo na icona

Favorito

A predición que se visualizará por defecto nesta páxina (e tamén na caixa de Concellos da página principal) será a do concello marcado con . Pode cambiar esta opción premendo sobre o botón do concello que queira seleccionar como favorito.

### Santiago de Compostela

**Predición a curto prazo**

DATA DE PREDICIÓN	xoves, 18 de xullo			venres, 19 de xullo			sábado, 20 de xullo			domingo, 21 de xullo						
	FRANXA HORARIA	Mañá	Tarde	Noite	Mañá	Tarde	Noite	Mañá	Tarde	Noite	Mañá	Tarde	Noite			
Estado do ceo																
Vento																
Probabilidade de choiva	5%	5%	5%	5%	15%	5%	5%	40%	20%	35%	5%	5%				
Temperaturas (°C)	MIN	MAX	14°	33°	MIN	MAX	16°	31°	MIN	MAX	15°	29°	MIN	MAX	15°	24°
Data de actualización	18/07/2013 09:00			18/07/2013 10:00			18/07/2013 11:00			18/07/2013 11:00						

Orixé dos datos: As previsións amosadas neste apartado son elaboradas polo equipo de Predición Operativa de MeteoGalicia.

**Predición a medio prazo para Santiago de Compostela**

DATA DE PREDICIÓN	luns, 22 de xullo		martes, 23 de xullo		mércores, 24 de xullo		xoves, 25 de xullo		venres, 26 de xullo							
	Estado do ceo	Vento	Estado do ceo	Vento	Estado do ceo	Vento	Estado do ceo	Vento	Estado do ceo	Vento						
Estado do ceo																
Vento	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR	VAR							
Probabilidade de choiva	5%		15%		30%		35%		70%							
Temperaturas (°C)	MIN	MAX	16°	28°	MIN	MAX	16°	28°	MIN	MAX	15°	26°	MIN	MAX	15°	25°
Comentario para Galicia	Durante a vinda semana Galicia atoparase nunha situación intermedia entre o anticíon dos Azores que non conseguirá estabilizar de novo a situación ata o final da semana e unha zona de baixas presións relativas e aire frío en altura algo máis ao norte. Isto manterá o tempo algo inseguro entre o luns e o venres. Agardamos en xeral un aumento das nubes, que ocasionalmente poderían deixar algúna chuvia feble e un descenso de temperatura, algo máis notorio no litoral, onde por momentos as temperaturas serán baixas para a época do ano.															
Data de actualización	18/07/2013 13:00															

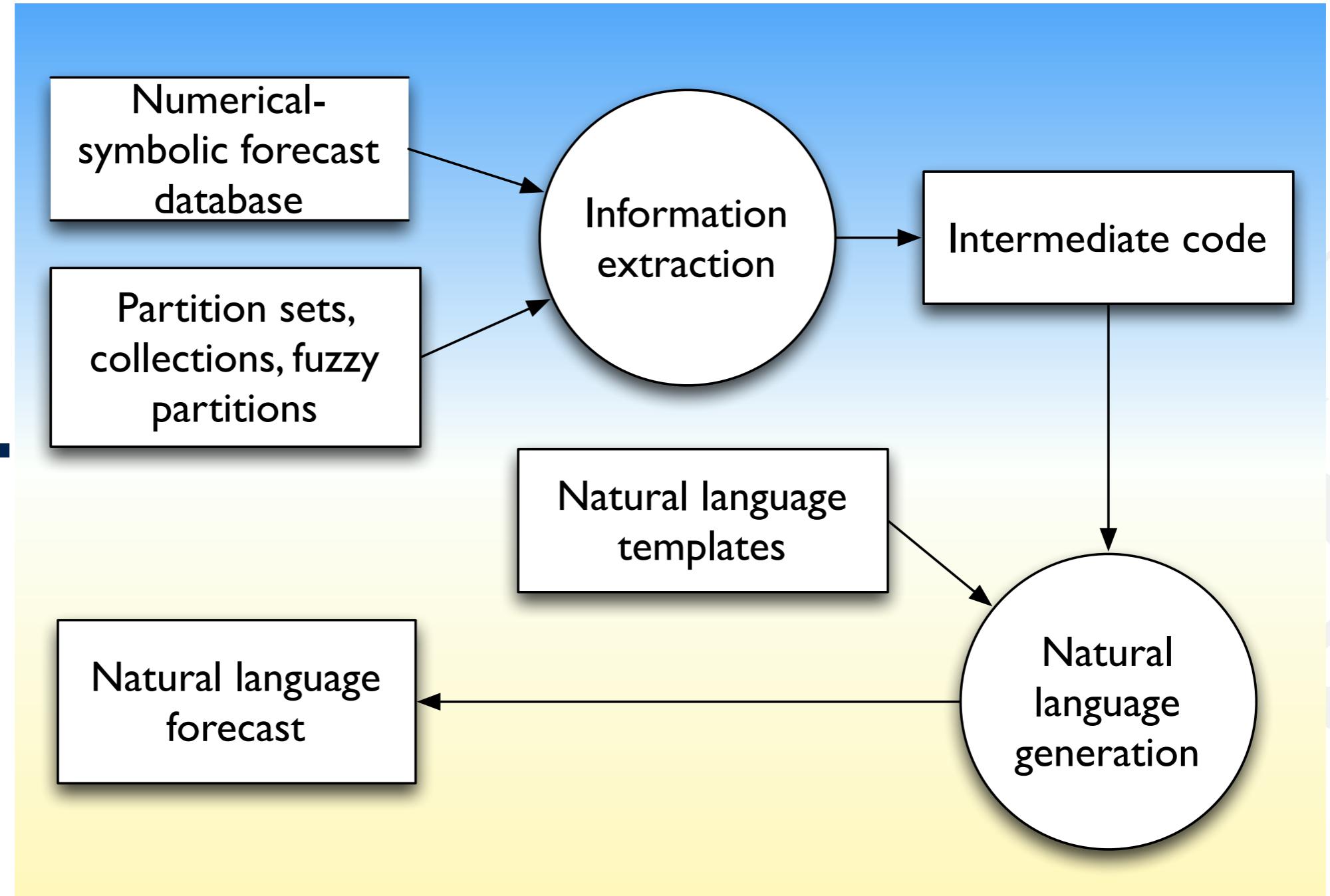
# Short-term weather forecasts

## Our approach

Linguistic  
description  
stage



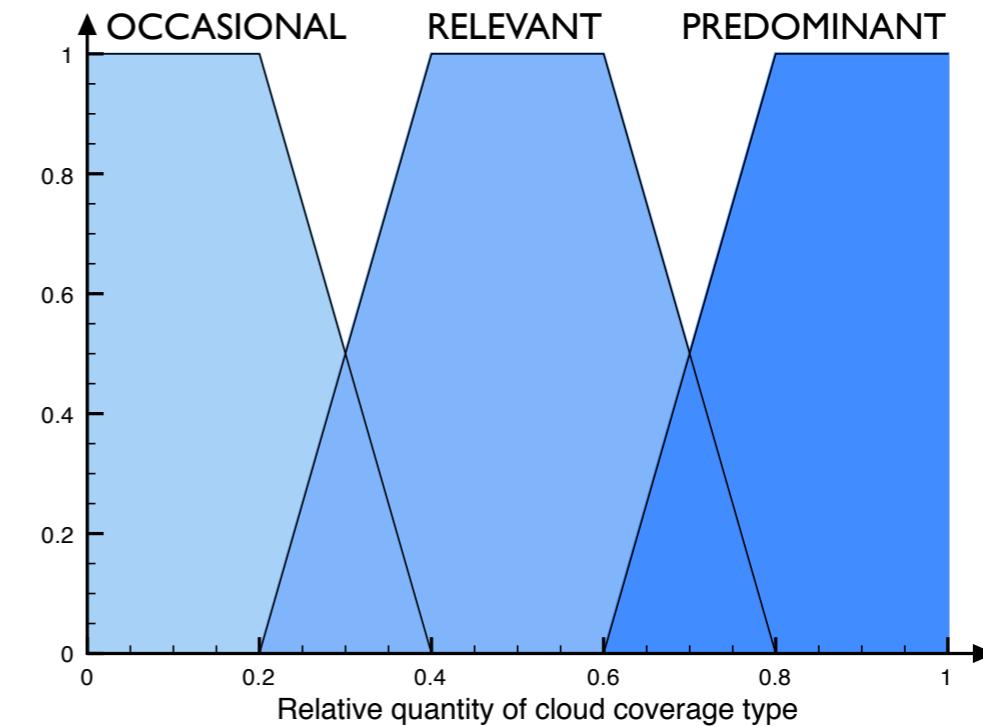
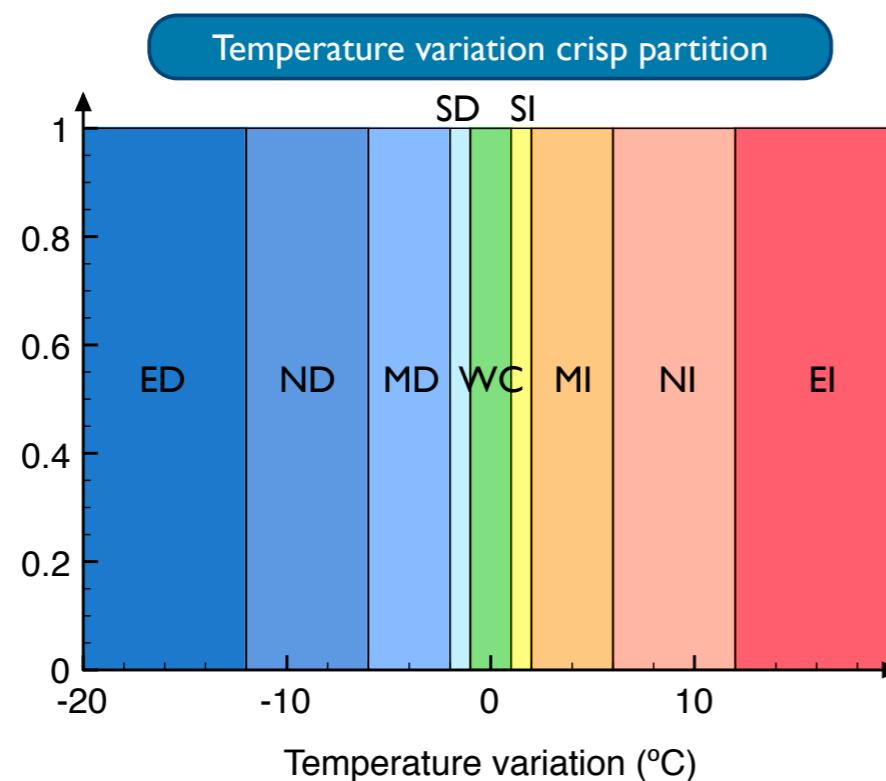
Natural  
language  
stage



# Short-term weather forecasts

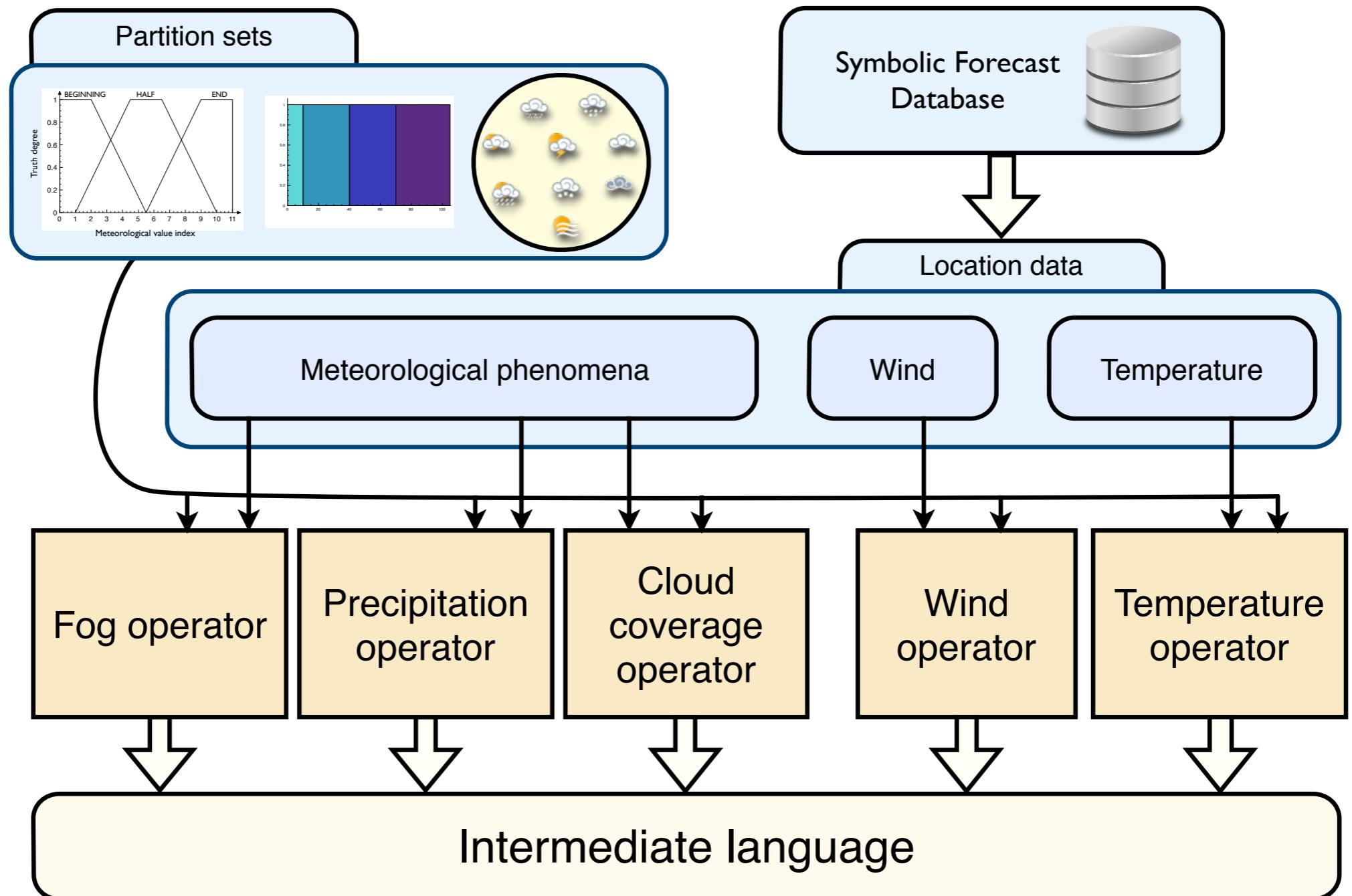
## Linguistic description extraction stage

- Operators extract the relevant information from the source data
- The extracted information is encoded into linguistic labels and time references, using an intermediate language
- In order to obtain relevant information from the source data we use symbol sets and crisp and fuzzy partitions



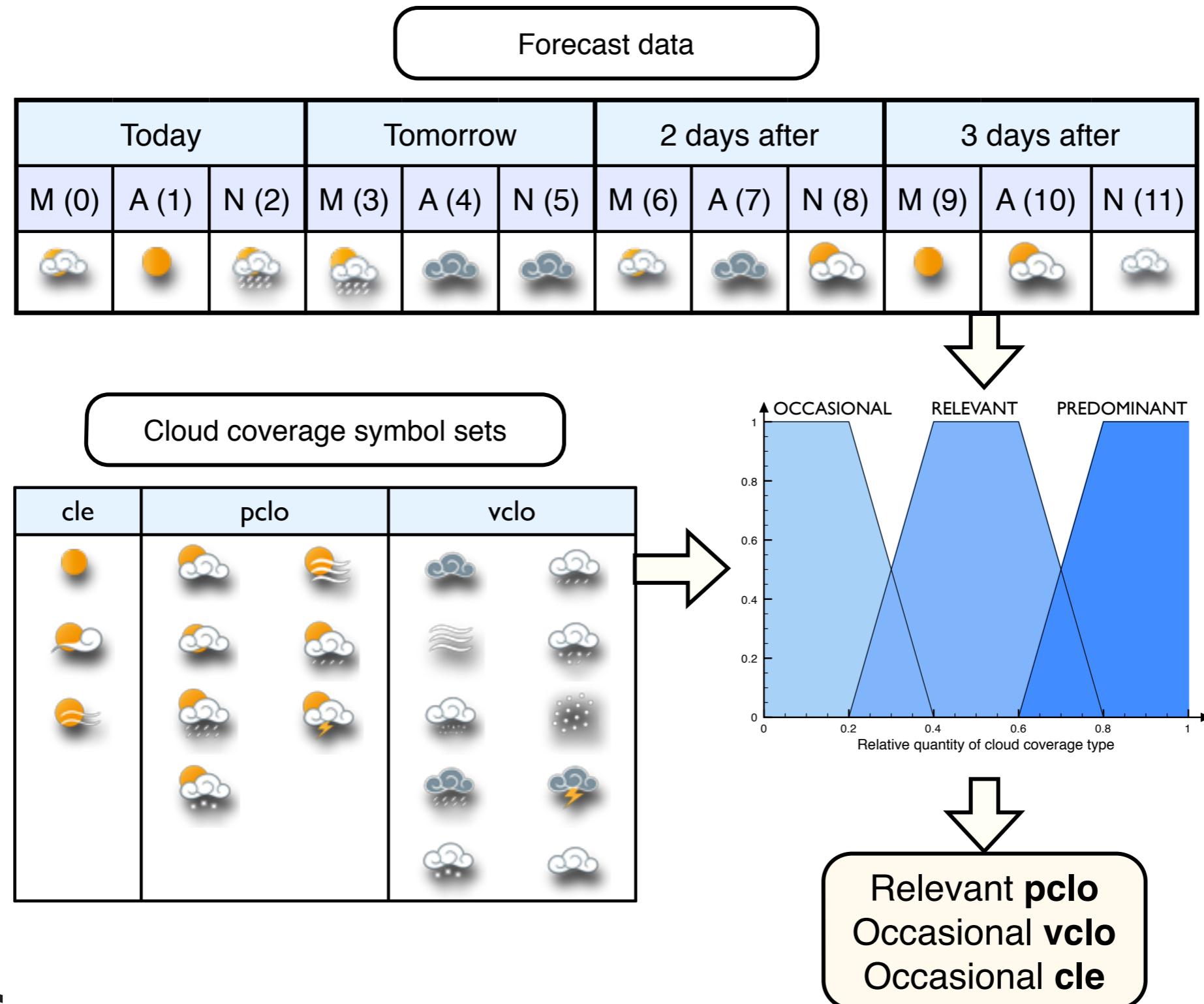
# Short-term weather forecasts

## Linguistic description extraction stage



# Short-term weather forecasts

Linguistic description extraction stage: Cloud coverage operator



# Short-term weather forecasts

## Natural language generation stage

- By using simple grammars the intermediate codes are translated into natural language texts
- We use XML templates to store the texts and generic expressions
- Two different natural language approaches:
  - Static templates:
    - Define the forecast text structures for certain variables
    - They contain expressions and labels used in both NLG approaches
  - Natural language generation logic:
    - For variables whose forecast textual structure is not static
    - Based on ideas from standard NLG approaches

# Short-term weather forecasts

## Natural language generation stage

```
<template id="T" name="Temperature">
  <case id="1">
    <part>
      <static>Las temperaturas</static>
      <space/>
      <static>serán</static>
      <space/>
      <option id="1">
        <variable id="norT"></variable>
        <space/>
        <static>para esta época del año,</static>
      </option>
      <option id="2">
        <variable id="minT"></variable>
        <space/>
        <static>para las mínimas y</static>
        <space/>
        <variable id="maxT"></variable>
        <space/>
        <static>para las máximas respecto a lo ha</static>
      </option>
      <space/>
    </part>
```

```
<labelset name="C" value="Cloud Coverage">
  <label name="C">poco nubosos o despejados</label>
  <label name="CL">parcialmente nubosas</label>
  <label name="O">muy nubosas</label>
  <label name="F">con nieblas</label>
</labelset>
<labelset name="R" value="Rain">
  <label name="I">intermitentes</label>
  <label name="P">persistentes</label>
  <label name="SN">de nieve</label>
  <label name="ST">tormentosas</label>
  <label name="H">de granizo</label>
</labelset>
<labelset name="V" value="Temperature Variation">
  <label name="ED">en descenso extremo</label>
  <label name="ND">en descenso notable</label>
  <label name="MD">en descenso moderado</label>
  <label name="SD">en descenso ligero</label>
  <label name="WC">sin cambios</label>
  <label name="SI">en ascenso ligero</label>
  <label name="MI">en ascenso moderado</label>
  <label name="NI">en ascenso notable</label>
  <label name="EI">en ascenso extremo</label>
</labelset>
```

# Short-term weather forecasts

## Examples

lunes			martes			miércoles			jueves		
mañana	tarde	noche	mañana	tarde	noche	mañana	tarde	noche	mañana	tarde	noche
Min: 5° - Max: 8°	Min: 2° - Max: 5°	Min: 0° - Max: 3°	Min: 2° - Max: 4°								

Cielos parcialmente nubosos en general durante los próximos días, aunque ocasionalmente se encontrarán poco nubosos o despejados. Tendremos precipitaciones todos los días, que podrán ser de nieve el miércoles por la mañana. Las temperaturas serán normales para las mínimas y muy bajas para las máximas respecto a lo habitual en esta época del año, con termómetros que globalmente se encontrarán en descenso moderado aunque oscilarán.

# Short-term weather forecasts

## Examples

lunes			martes			miércoles			jueves		
mañana	tarde	noche	mañana	tarde	noche	mañana	tarde	noche	mañana	tarde	noche
Min: 10° - Max: 15°	Min: 11° - Max: 16°	Min: 12° - Max: 18°	Min: 11° - Max: 16°								

Se espera una alternancia de periodos de cielos muy nubosos con otros parcialmente nubosos para los próximos días, aunque ocasionalmente podrán encontrarse poco nubosos o despejados. Habrá nieblas matinales el jueves (persistentes). Tendremos precipitaciones el martes y el miércoles, que podrán ser tormentosas el miércoles por la tarde. Las temperaturas serán muy altas para las mínimas y altas para las máximas respecto a lo habitual en esta época del año, con termómetros que globalmente se encontrarán sin cambios aunque oscilarán. Viento fuerte del Sudoeste desde el miércoles por la mañana hasta el miércoles por la tarde.

# Short-term weather forecasts

## Validation process

- The validation process of a linguistic description approach is vital
- We need to make sure that the texts our application generates are good enough to be released to the public
- Human expert assessment is the most reliable validation approach and it is widely accepted
- In our case, we have been fully supported by a meteorologist: Juan Taboada from MeteoGalicia



# Short-term weather forecasts

## Validation process

- We designed a questionnaire which covers several quality dimensions of the automatically generated natural language forecasts
- From a dataset comprised of 45 forecasts, which contain both synthetic and real cases, a set of respective 45 natural language forecasts were generated
- Our expert assigned a score to each textual forecast following our questionnaire

Average global score (out of 5)	Standard deviation
<b>4.77</b>	<b>0.18</b>

# Short-term weather forecasts

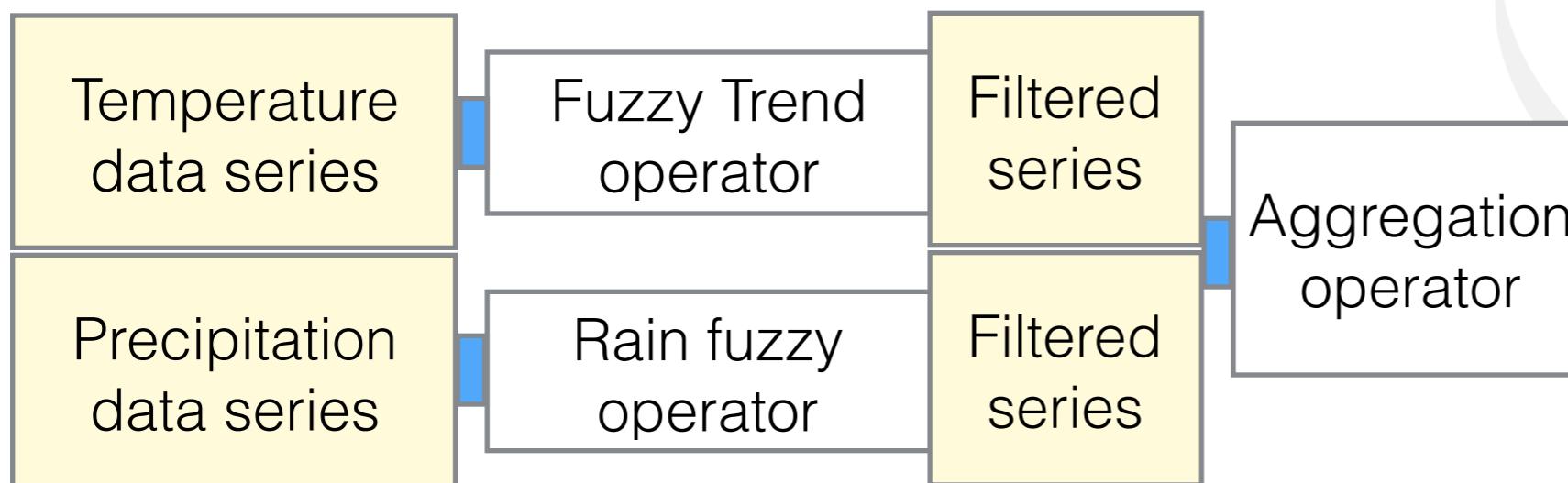
## Future work

- Our short-term forecast application will be transferred to MeteoGalicia in the coming weeks to start providing automatically generated natural language forecasts, which will be displayed in the municipalities section of their website
- We will extend the current forecasts with linguistic descriptions of the air quality
- Another future application field will reside in the project management domain
- We will explore more application fields in which linguistic descriptions prove useful
- Last but not least, we plan to design and develop a general linguistic description framework

# Future work

## Generic linguistic framework

- This framework would comprise a set of data structures, operators and objective quality criteria, which could be composed in order to obtain complex descriptions
- Its aim would be to easily provide linguistic description solutions for any application field
- Natural language generation modules would then be specifically developed for each linguistic description solution



Shortly after the high decrease in temperatures in the first week, strong precipitations took place during the second and third weeks

**Thank you for your attention  
Gracias por vuestra atención  
Grazas pola vosa atención**

**[alejandro.ramos@usc.es](mailto:alejandro.ramos@usc.es)**

**[citus.usc.es](http://citus.usc.es)**