

DIC and other model selection criteria in risk mapping

Ioana Molnar¹

in collaboration with Sylvain Coly^{1,2}, Myriam Charras-Garrido¹,
David Abrial¹ et Anne-Françoise Yao-Lafourcade²

1 : Unité d'Épidémiologie Animale (EPIA), Centre INRA Auvergne – Rhône-Alpes
2 : Laboratoire de Mathématiques, Université Blaise Pascal, Clermont-Ferrand



48^{èmes} Journées de Statistique

30 may – 3 june 2016

Montpellier, 31 may, 2016

Disease risk mapping

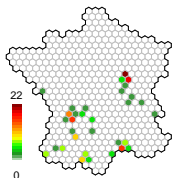
→ tool used in spatial statistics for the analysis of the **risk underlying the observed incidence** of a disease.

- Risk estimation: Bayesian methods

Disease risk mapping

→ tool used in spatial statistics for the analysis of the **risk underlying the observed incidence** of a disease.

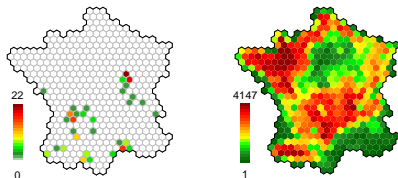
- Risk estimation: Bayesian methods
- A priori information:
 - case data
 - population data
 - parametric distribution of cases
 - risk structure



Disease risk mapping

→ tool used in spatial statistics for the analysis of the **risk underlying the observed incidence** of a disease.

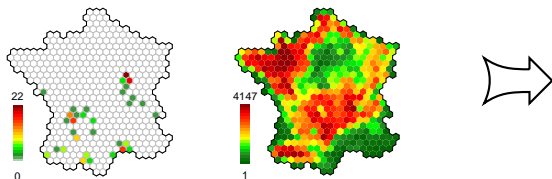
- Risk estimation: Bayesian methods
- A priori information:
 - case data
 - population data
 - parametric distribution of cases
 - risk structure



Disease risk mapping

→ tool used in spatial statistics for the analysis of the **risk underlying the observed incidence** of a disease.

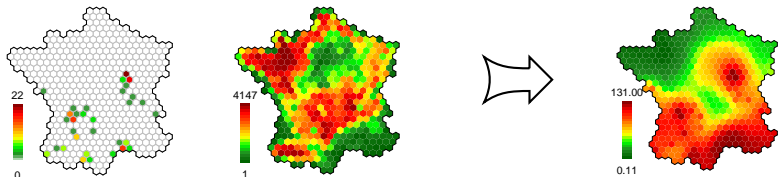
- Risk estimation: Bayesian methods
- A priori information:
 - case data
 - population data
 - parametric distribution of cases
 - risk structure



Disease risk mapping

→ tool used in spatial statistics for the analysis of the **risk underlying the observed incidence** of a disease.

- Risk estimation: Bayesian methods
- A priori information:
 - case data
 - population data
 - parametric distribution of cases
 - risk structure
- Risk representation: smooth maps



Model selection in risk mapping

- Properties of a good model?
 - fits well on the data
 - is parsimonious
 - performs data smoothing
 - has a good explicative (and predictive) power
- Properties of a good model selection criterion?
 - identifies the *best* models
 - filters out the unsuitable models
 - ranks the pertinent ones
 - easily implementable, easy-to-use
 - robust

Model selection in risk mapping

- Properties of a good model?
 - fits well on the data
 - is parsimonious
 - performs data smoothing
 - has a good explicative (and predictive) power
- Properties of a good model selection criterion?
 - identifies the *best* models
 - filters out the unsuitable models
 - ranks the pertinent ones
 - easily implementable, easy-to-use
 - robust

} frequently related

Model selection in risk mapping

- Properties of a good model?
 - fits well on the data
 - is parsimonious
 - performs data smoothing
 - has a good explicative (and predictive) power
- Properties of a good model selection criterion?
 - identifies the *best* models
 - filters out the unsuitable models
 - ranks the pertinent ones
 - easily implementable, easy-to-use
 - robust

} frequently related

The DIC in risk mapping

DIC = Deviance Information Criterion
(SPIEGELHALTER *et al.*, 2002)

→ based on the *Bayesian deviance* $D(\theta) := -2 \ln p(Y | \theta)$

→ measures the INADEQUACY penalized by the COMPLEXITY

↘ $\overline{D(\theta)}$

↘ $p_D := \overline{D(\theta)} - D(\bar{\theta})$

effectif # of parameters

Definition:

$$\text{DIC} := 2\overline{D(\theta)} - D(\bar{\theta}).$$

ADVANTAGES

SHORTCOMINGS

The DIC in risk mapping

DIC = Deviance Information Criterion
(SPIEGELHALTER *et al.*, 2002)

→ based on the *Bayesian deviance* $D(\theta) := -2 \ln p(Y | \theta)$

→ measures the INADEQUACY penalized by the COMPLEXITY

↘ $\overline{D(\theta)}$

↘ $p_D := \overline{D(\theta)} - D(\bar{\theta})$

effectif # of parameters

Definition:

$$DIC := 2\overline{D(\theta)} - D(\bar{\theta}).$$

ADVANTAGES

- No need to compute the number of parameters
- Easy computation using MCMC samples
- Lack of a strong competitor

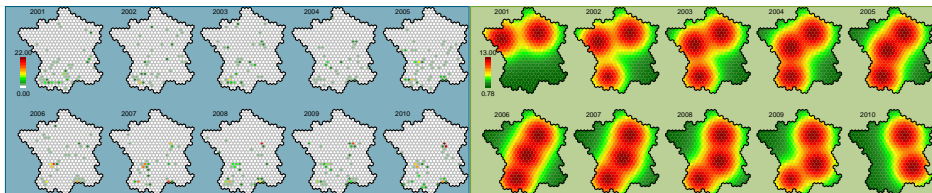
SHORTCOMINGS

- p_D sometimes negative
- Favors overfitted models
- Lacks invariance to re-parametrization, and others

Context of study of the DIC: the data

Two types of data:

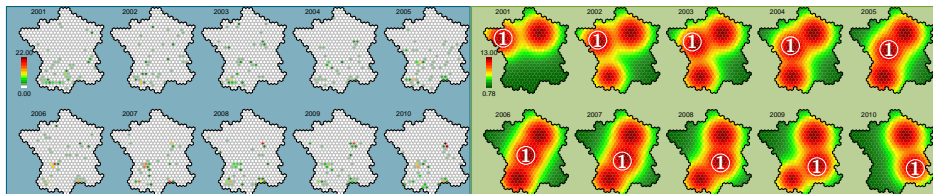
- ↗ real data of bovine tuberculosis in France: overdispersed
- ↘ simulated data from a known risk: by \mathcal{BN}



Context of study of the DIC: the data

Two types of data:

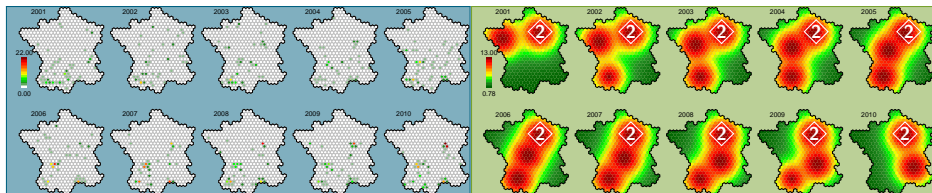
- ↗ real data of bovine tuberculosis in France: overdispersed
- ↘ simulated data from a known risk: by \mathcal{BN}



Context of study of the DIC: the data

Two types of data:

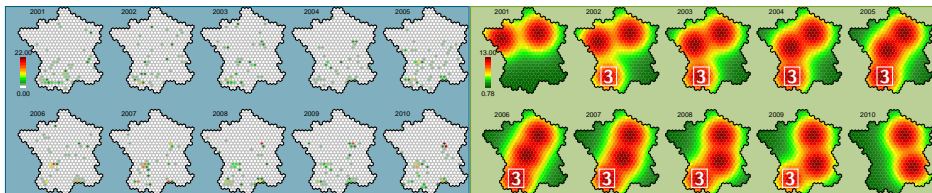
- ↗ real data of bovine tuberculosis in France: overdispersed
- ↘ simulated data from a known risk: by \mathcal{BN}



Context of study of the DIC: the data

Two types of data:

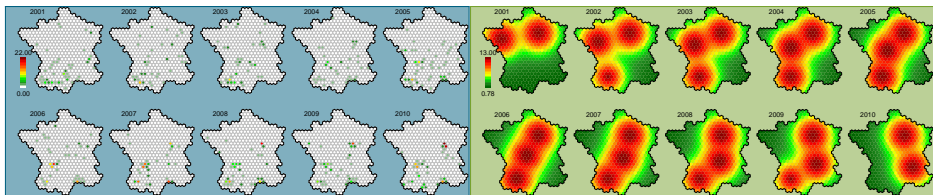
- ↗ real data of bovine tuberculosis in France: overdispersed
- ↘ simulated data from a known risk: by \mathcal{BN}



Context of study of the DIC: the data

Two types of data:

- ↗ real data of bovine tuberculosis in France: overdispersed
- ↘ simulated data from a known risk: by \mathcal{BN}



Goal: select the models for which the risk maps:

- fit the data
- are smooth enough – consistence with Moran's I?
- show well-delimited structures
- are close to the real underlying risk – consistence with MSE / Spearman's ρ ?

Context of study of the DIC: the models

→ spatio-temporal, three-level hierarchical Bayesian models

Data: $Y_i^j = \#$ of cases in region i and period j
 $i = 1, \dots, n := 448, \quad j = 1, \dots, m := 10.$

Modeling:

- *1st Level* – the distribution of the cases:

$$Y_i^j \sim \mathcal{P}(\lambda_i^j) \quad \text{or} \quad \begin{cases} Y_i^j \sim \mathcal{P}(\lambda_i^j) \\ \lambda_i^j \sim \gamma(\alpha_i^j, \beta_i^j) \end{cases}$$

Parameter of interest: $\lambda_i^j \approx$ relative risk R_i^j .

- *2nd Level* – relative risk structure:

$$\left\{ \begin{array}{l} \ln R_i^j = \underbrace{\alpha \cdot U_i^j}_{\text{spatial component}} + \underbrace{b \cdot T_i^j}_{\text{temporal component}} + \underbrace{c \cdot V_i^j}_{\text{spatio-temporal component}} + \underbrace{d \cdot \varepsilon_i^j}_{\text{white noise}} \\ U_i^j, T_i^j, V_i^j \rightarrow \text{CAR-type process} \\ \alpha, b, c, d \rightarrow \text{weights: } \alpha, b, c \in \{0, 1\} \quad \text{or} \quad \alpha, b, c \sim \gamma(5, 5), d \in \{0, 1\}. \end{array} \right.$$

- *3rd Level* – specification of the precision parameters of the normal distributions of $\varepsilon_i^j, U_i^j, T_i^j, V_i^j$: $\gamma(0.01, 0.01)$.

Context of study of the DIC: the models

→ spatio-temporal, three-level hierarchical Bayesian models

Data: $Y_i^j = \#$ of cases in region i and period j
 $i = 1, \dots, n := 448, \quad j = 1, \dots, m := 10.$

Modeling:

- *1st Level* – the distribution of the cases:

$$Y_i^j \sim \mathcal{P}(\lambda_i^j) \quad \text{or} \quad \begin{cases} Y_i^j \sim \mathcal{P}(\lambda_i^j) \\ \lambda_i^j \sim \gamma(\alpha_i^j, \beta_i^j) \end{cases}$$

Parameter of interest: $\lambda_i^j \approx$ relative risk R_i^j .

- *2nd Level* – relative risk structure:

$$\left\{ \begin{array}{l} \ln R_i^j = \underbrace{a \cdot U_i^j}_{\text{spatial component}} + \underbrace{b \cdot T_i^j}_{\text{temporal component}} + \underbrace{c \cdot V_i^j}_{\text{spatio-temporal component}} + \underbrace{d \cdot \varepsilon_i^j}_{\text{white noise}} \\ U_i^j, T_i^j, V_i^j \rightarrow \text{CAR-type process} \\ a, b, c, d \rightarrow \text{weights: } a, b, c \in \{0, 1\} \quad \text{or} \quad a, b, c \sim \gamma(5, 5), d \in \{0, 1\}. \end{array} \right.$$

- *3rd Level* – specification of the precision parameters of the normal distributions of $\varepsilon_i^j, U_i^j, T_i^j, V_i^j$: $\gamma(0.01, 0.01)$.

Context of study of the DIC: the models

→ spatio-temporal, three-level hierarchical Bayesian models

Data: $Y_i^j = \#$ of cases in region i and period j
 $i = 1, \dots, n := 448, \quad j = 1, \dots, m := 10.$

Modeling:

- *1st Level* – the distribution of the cases:

$$Y_i^j \sim \mathcal{P}(\lambda_i^j) \quad \text{or} \quad \begin{cases} Y_i^j \sim \mathcal{P}(\lambda_i^j) \\ \lambda_i^j \sim \gamma(\alpha_i^j, \beta_i^j) \end{cases}$$

Parameter of interest: $\lambda_i^j \approx$ relative risk R_i^j .

- *2nd Level* – relative risk structure:

$$\left\{ \begin{array}{l} \ln R_i^j = \underbrace{\alpha \cdot U_i^j}_{\text{spatial component}} + \underbrace{b \cdot T_i^j}_{\text{temporal component}} + \underbrace{c \cdot V_i^j}_{\text{spatio-temporal component}} + \underbrace{d \cdot \varepsilon_i^j}_{\text{white noise}} \\ U_i^j, T_i^j, V_i^j \rightarrow \text{CAR-type process} \\ \alpha, b, c, d \rightarrow \text{weights: } \alpha, b, c \in \{0, 1\} \quad \text{or} \quad \alpha, b, c \sim \gamma(5, 5), d \in \{0, 1\}. \end{array} \right.$$

- *3rd Level* – specification of the precision parameters of the normal distributions of $\varepsilon_i^j, U_i^j, T_i^j, V_i^j$: $\gamma(0.01, 0.01)$.

Context of study of the DIC: the models

→ spatio-temporal hierarchical Bayesian models

Data: $Y_i^j = \#$ of cases in region i and period j
 $i = 1, \dots, n := 448, j = 1, \dots, m := 10.$

Model:

Parameter of interest: $\lambda_j^i \approx$ relative risk R_{j+1}^i

- 2nd Level – relative risk structure:

$$\left\{ \begin{array}{l} \ln R_i^j = \underbrace{a \cdot U_i^j}_{\text{spatial component}} + \underbrace{b \cdot T_i^j}_{\text{temporal component}} + \underbrace{c \cdot V_i^j}_{\text{spatio-temporal component}} + \underbrace{d \cdot \varepsilon_i^j}_{\text{white noise}} \\ U_i^j, T_i^j, V_i^j \rightarrow \text{CAR-type process} \\ a, b, c, d \rightarrow \text{weights: } a, b, c \in \{0, 1\} \text{ or } a, b, c \sim \gamma(5, 5), d \in \{0, 1\}. \end{array} \right.$$

- 3rd Level – specification of the precision parameters of the normal distributions of $\varepsilon_i^j, U_i^j, T_i^j, V_i^j$: $\gamma(0.01, 0.01)$.

Context of study of the DIC: the models

→ spatio-temporal, three-level hierarchical Bayesian models

Data: $Y_i^j = \#$ of cases in region i and period j
 $i = 1, \dots, n := 448, \quad j = 1, \dots, m := 10.$

Modeling:

- *1st Level* – the distribution of the cases:

$$Y_i^j \sim \mathcal{P}(\lambda_i^j) \quad \text{or} \quad \begin{cases} Y_i^j \sim \mathcal{P}(\lambda_i^j) \\ \lambda_i^j \sim \gamma(\alpha_i^j, \beta_i^j) \end{cases}$$

Parameter of interest: $\lambda_i^j \approx$ relative risk R_i^j .

- *2nd Level* – relative risk structure:

$$\begin{cases} \ln R_i^j = \underbrace{\alpha \cdot U_i^j}_{\text{spatial component}} + \underbrace{\beta \cdot T_i^j}_{\text{temporal component}} + \underbrace{c \cdot V_i^j}_{\text{spatio-temporal component}} + \underbrace{d \cdot \varepsilon_i^j}_{\text{white noise}} \\ U_i^j, T_i^j, V_i^j \rightarrow \text{CAR-type process} \\ \alpha, \beta, c, d \rightarrow \text{weights: } \alpha, \beta, c \in \{0, 1\} \quad \text{or} \quad \alpha, \beta, c \sim \gamma(5, 5), d \in \{0, 1\}. \end{cases}$$

- *3rd Level* – specification of the precision parameters of the normal distributions of $\varepsilon_i^j, U_i^j, T_i^j, V_i^j$: $\gamma(0.01, 0.01)$.

Estimation of the parameters

$$\begin{cases} \lambda_i^j \sim \mathcal{P} \text{ or } \lambda_i^j \sim \mathcal{BN} \\ \ln R_i^j = aU_i^j + bV_i^j + cT_i^j + \varepsilon_i^j \end{cases}$$

of models: 60, depending on the choice of

→ the 1st level distribution: \mathcal{P} or \mathcal{BN} .

→ the random effects included aU, bT, cV and / or ε .

of data sets: 1 (real data); 100 (simulated data).

of estimation replications: 100 (real data); 1 (simulated data).

of parameters: up to $\approx 22\,400$, depending on the model.

Estimation: by MCMC under OpenBUGS.

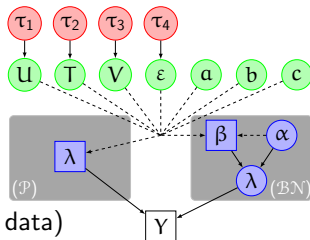
BUGS parameters:

random seed: from 1 to 14 (real data)

burn-in step: from 10 000 to 80 000 (real data)

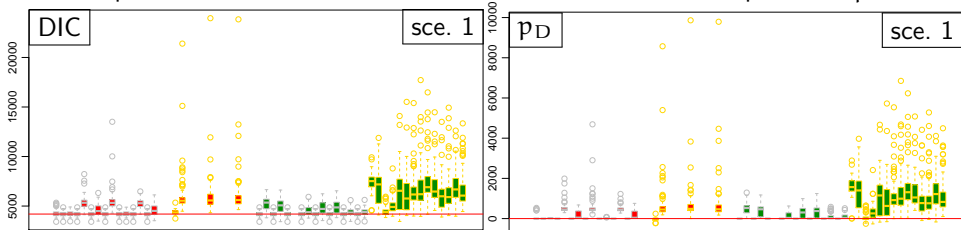
thinning step: 10

sample length: 10 000.



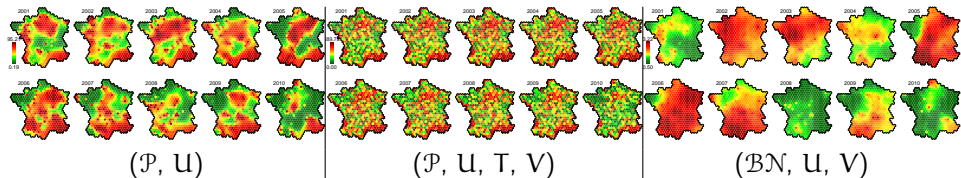
The DIC for the simulated data

Boxplots of the values of the DIC and of the effective number of parameters p_D :



Comments:

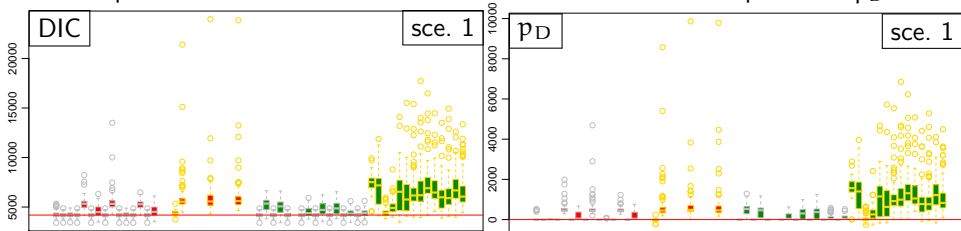
→ Identical values for 16 of the models (those with no weights nor noise).



→ Negative values for p_D for several models, null values for many others.

The DIC for the simulated data

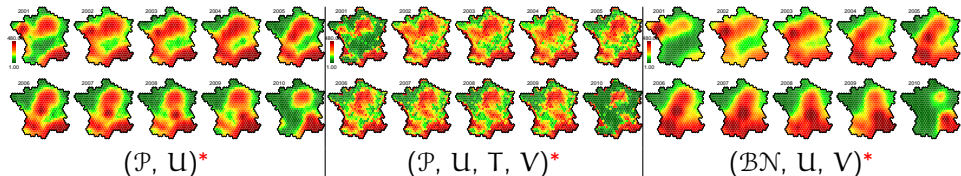
Boxplots of the values of the DIC and of the effective number of parameters p_D :



Comments:

→ Identical values for 16 of the models (those with no weights nor noise).

mean over replicats



→ Negative values for p_D for several models, null values for many others.

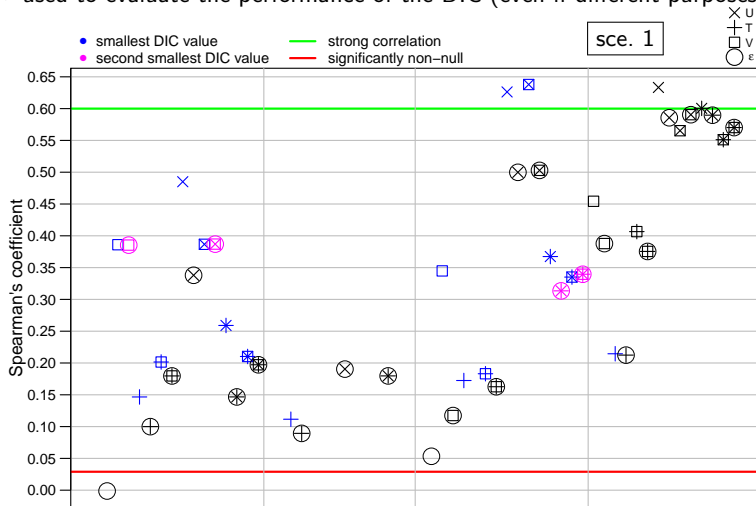
Other classical comparison methods

- Accuracy measures (MSE related)
- Spearman's rank correlation coefficient ρ
 - applied on the relative risk (not the count data)
 - may be used only on simulation studies
 - measure the adequacy of the model
 - miss the notion of parsimony
- Spatio-temporal association indicators (Moran's I, Geary's c)
 - standard spatial association indices, extended to the spatio-temporal context
 - may be used on real data
 - evaluate the smoothness of the risk maps
 - lack the idea of adjustment
- * Naturally, they won't favor the same models as the DIC (they have **different** purposes)
- * However, we can use them to evaluate **some aspects** of the DIC's performance in choosing "good" models.
- * They can be used to set up a better selection tool.

The DIC vs. Spearman's ρ (simulated data)

Spearman's ρ :

→ used to evaluate the performance of the DIC (even if different purposes)

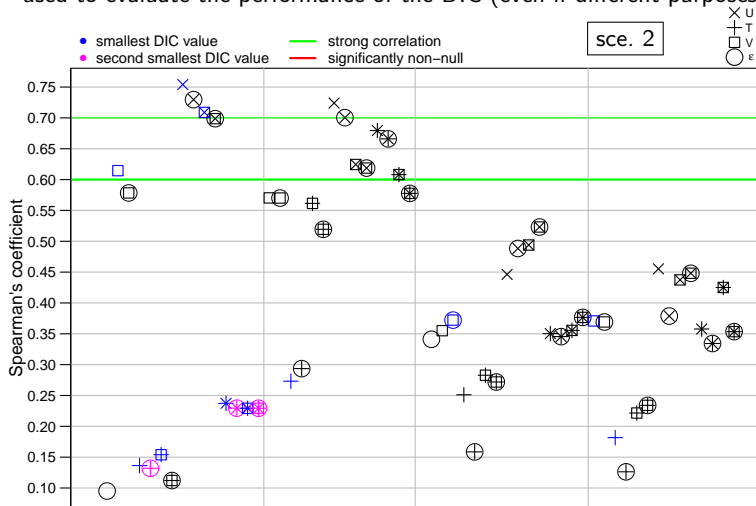


* Models with good DIC values show highly variable ρ 's.

The DIC vs. Spearman's ρ (simulated data)

Spearman's ρ :

→ used to evaluate the performance of the DIC (even if different purposes)



* Models with good DIC values show highly variable ρ 's.

Proposition of new selection methods

- Criteria coupling: $C(C_1, C_2)$
 - Filters out models that are “bad” according to C_1 , and selects one good model according to C_2
 - Order matters
 - Easy to apply, but needs some post-hoc analysis

- A new deviance criterion: Smoothness Deviance Criterion (SDC)

- Replaces the penalization by the **complexity** measure p_D with a penalization for the **lack of smoothness**

$$SDC = (\overline{D(\theta)} + \xi) \left(1 + \frac{1}{I \cdot \mathbb{1}_{[0,1]}} \right),$$

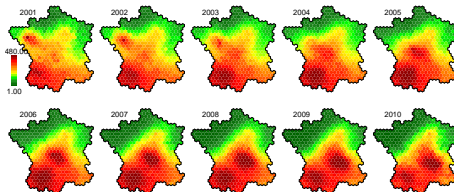
where $\xi > 0$ s.t. $\overline{D(\theta)} + \xi > 0$ for all compared models.

- Multiplicative penalization \implies the hierarchy defined by the goodness of fit measure $\overline{D(\theta)}$ is modified only if the models have very poor smoothness coefficient
- The parsimony idea is contained in I
- Easy to implement

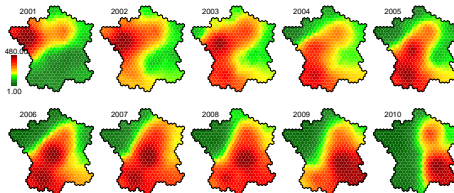
Application of the new methods on the simulated data

For both scenarios, the model selected *via* the coupling **C**(Moran's I, DIC) and that selected *via* the **SDC** are the same.

- For scenario 1 (data generated by \mathcal{BN}): model \mathcal{BN} , V .

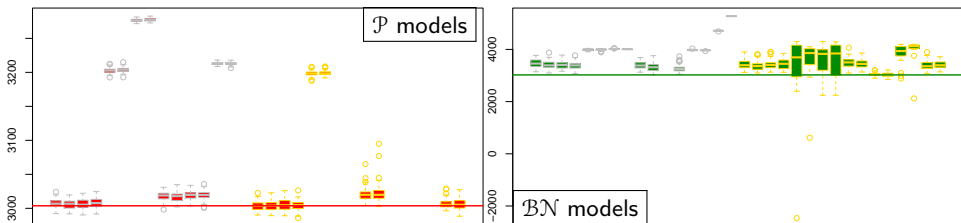


- For scenario 2 (data generated by \mathcal{P}): model \mathcal{P} , U .



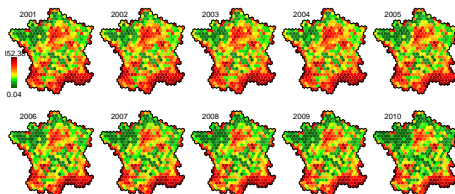
The DIC for the real data

Boxplots of the values of the DIC:



→ Smallest DIC values obtained for several models
(all of them \mathcal{P} models).

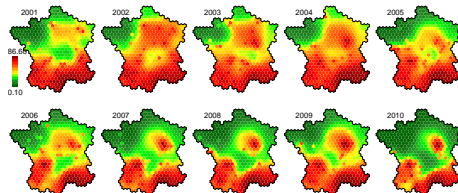
Example: $(\mathcal{P}, \mathcal{U}, \mathcal{T}, \mathcal{V})$.



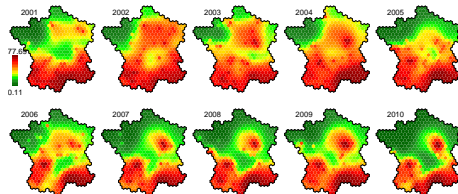
Application of the new methods on the real data

The models selected *via* the coupling **C(DIC, Moran's I)**, and that selected *via* the **SDIC** are different, but very similar.

- C(DIC, Moran's I): model (\mathcal{BN} , U , V).



- SDC: model (\mathcal{BN} , U).



Summary:

- Weaknesses of the DIC specific to disease mapping:
 - lack of smoothness of the risk maps
 - identical ranking of models producing very different risk maps
 - too much divergence between the DIC and the accuracy measures
- The systematic application of the DIC in disease mapping is not justified.
- The modification of the DIC to take into account the smoothness (crucial point in risk mapping) through *Criteria Coupling* and *SDC* shows promising results.

Some work to follow:

- confirm the utility of the proposed models on different data sets and simulation scenarios
- investigate the properties of the proposed method (robustness)

Summary:

- Weaknesses of the DIC specific to disease mapping:
 - lack of smoothness of the risk maps
 - identical ranking of models producing very different risk maps
 - too much divergence between the DIC and the accuracy measures
- The systematic application of the DIC in disease mapping is not justified.
- The modification of the DIC to take into account the smoothness (crucial point in risk mapping) through *Criteria Coupling* and *SDC* shows promising results.

Some work to follow:

- confirm the utility of the proposed models on different data sets and simulation scenarios
- investigate the properties of the proposed method (robustness)

Thank you!