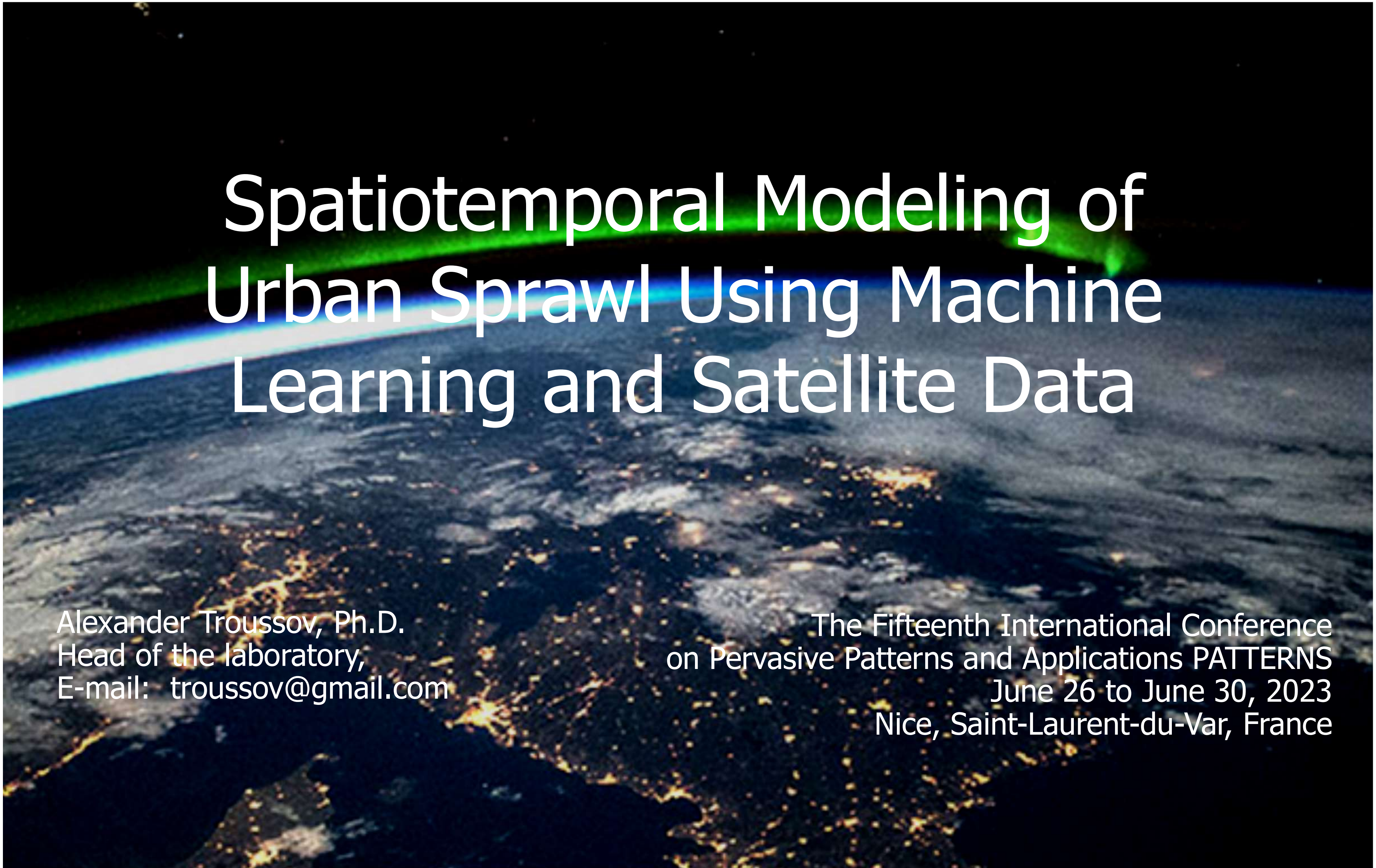
RANEPA

International laboratory
for mathematical modelling
of social networks



Spatiotemporal Modeling of Urban Sprawl Using Machine Learning and Satellite Data

Alexander Troussov, Ph.D.
Head of the laboratory,
E-mail: troussov@gmail.com

The Fifteenth International Conference
on Pervasive Patterns and Applications PATTERNS
June 26 to June 30, 2023
Nice, Saint-Laurent-du-Var, France

ALEXANDER TROUSSOV (aka Alexander Trusov)

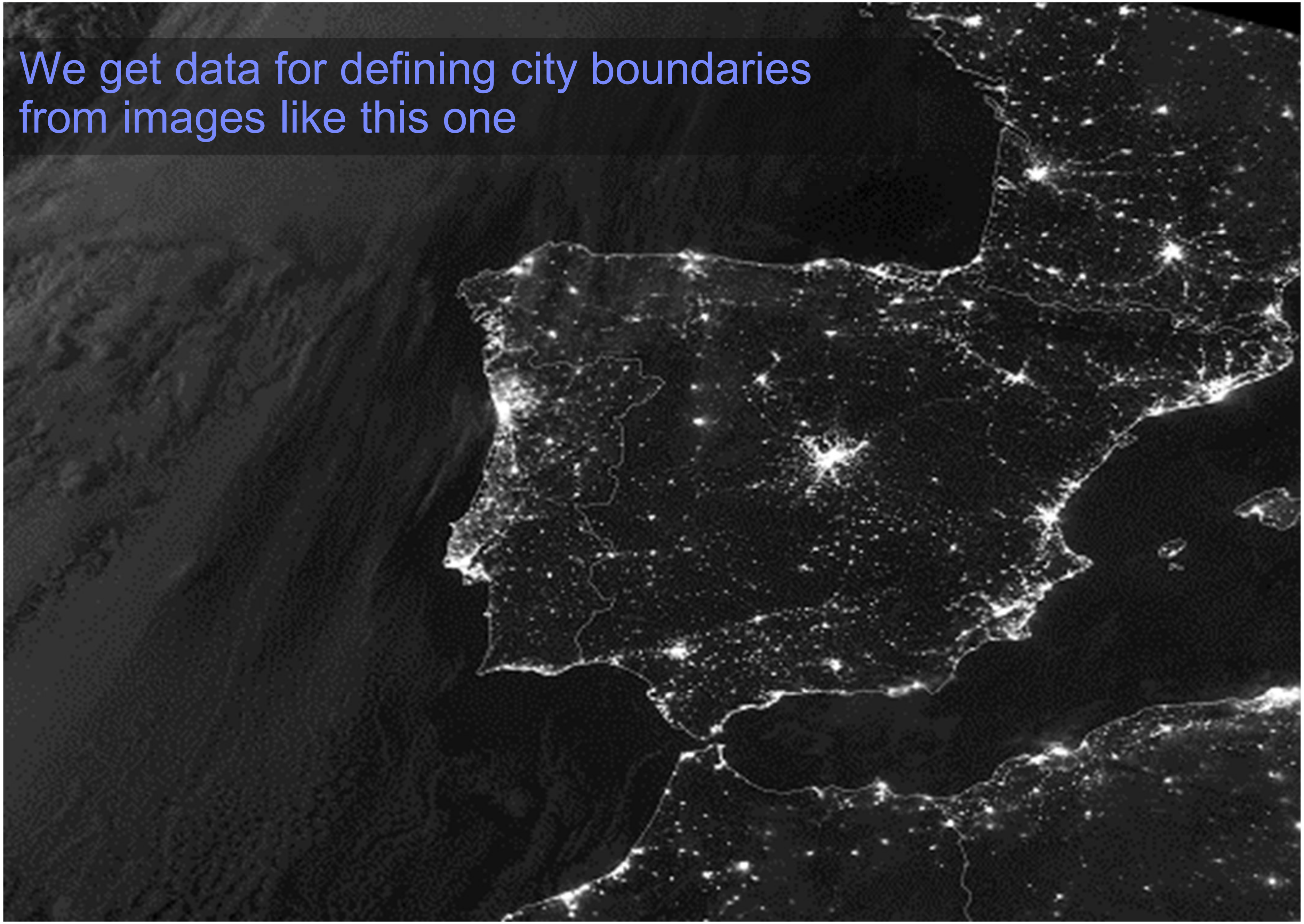
- 2014-   : Director of the International Research Laboratory for Mathematical Methods for Social Network Mining at RANEPA
- 2000-2013: IBM
  - Architect of IBM LanguageWare
  - Chief Scientist of IBM Dublin Center for Advanced Studies
  - EU Projects
- National Geophysical Data Center, Boulder, CO, USA - Visiting scientist
  - Earth Remote Sensing, numerical databases
- Observatoire de la Côte d'Azur, Nice, France – Visiting scientist
  - Numerical simulation of turbulence
- Institute of Physics of the Earth (Russian Academy of Sciences) and the International Institute for Earthquake Prediction Theory and Mathematical Geophysics, Moscow, Russia – Lead Researcher
  - R&D in geophysics and geoinformatics
- Ph.D. in Pure mathematics from Lomonosov Moscow State University

Chernobyl 1999

EU-UNESCO workshop, organised within the EU funded project for Support for Telematics Applications Cooperation with the Commonwealth of Independent States (STACCIS)

In this photo you can see Alexander Troussov (STACCIS telematics application engineer, first from the left), Jean Bonnin (Professor of Louis Pasteur University in Strasbourg) and other scientists near the sarcophagus covering the destroyed Chernobyl nuclear reactor.

AN EXAMPLE OF NTL DATA



Iberian peninsula: The VIIRS DNB layer is created using a sensing technique designed to capture low-light emissions under varying illumination conditions, and is displayed as a gray-scale image

1

Stable Night-Time Lights (2015) - a composite image of stable NTL composed of many images cleared from noise and clutter.

What is a city (conurbation, megalopolis)?

Cities consume 70% of energy, provide 90% of GDP,

Defining the natural boundaries of urban-type settlements is an important task. But administrative borders may not coincide with the borders of places where people live compactly. There are a lot of visitors in cities, the residential population can be very different from the so-called ambient population.

The use of the "light footprint" of cities helps in determining the boundaries of compact living.

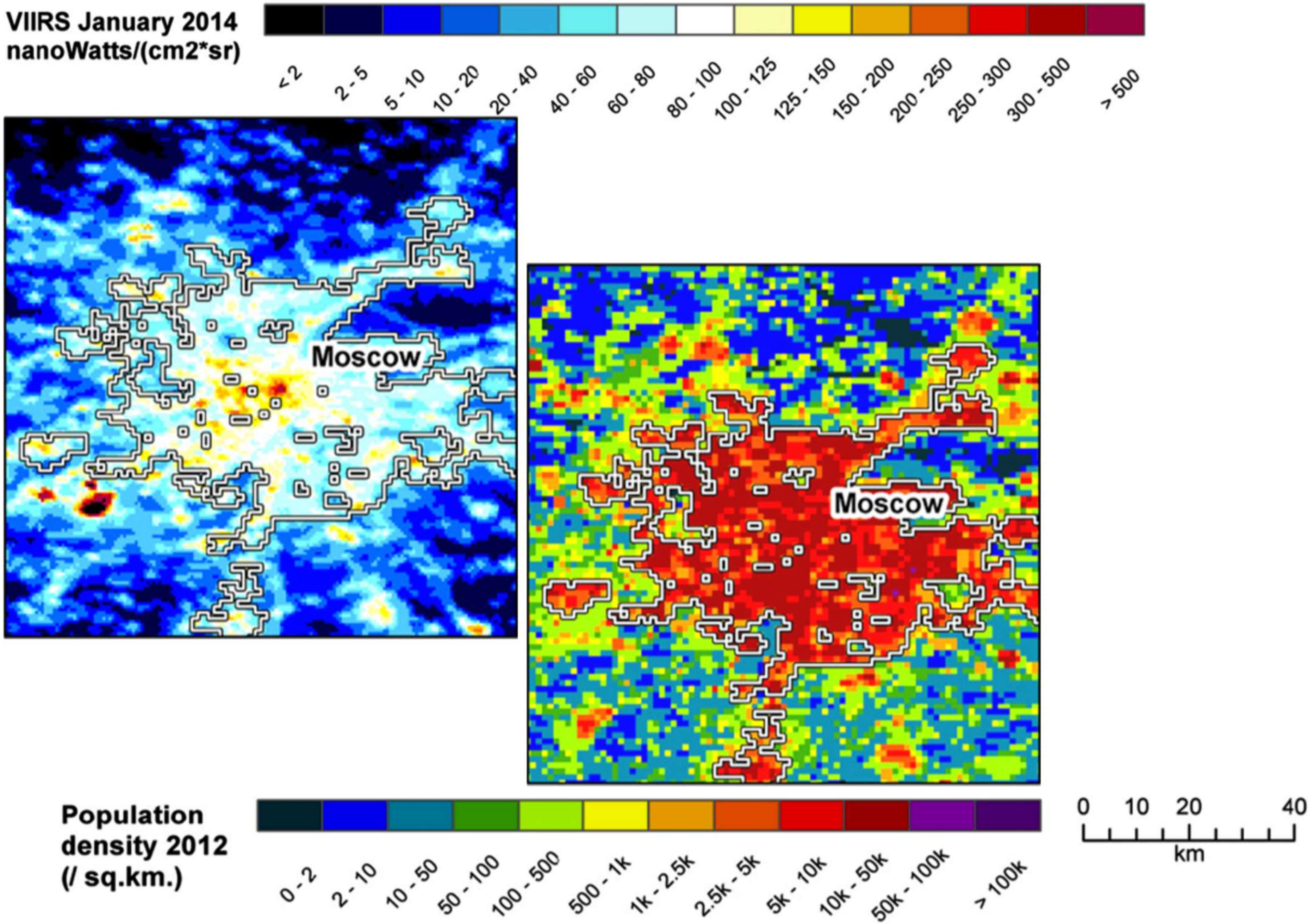
The idea is simple – to choose a suitable threshold of "brightness" - to call the entire territory where the lighting exceeds this threshold a city.

So, we have set ourselves two tasks in our project

1. separate the light from the darkness

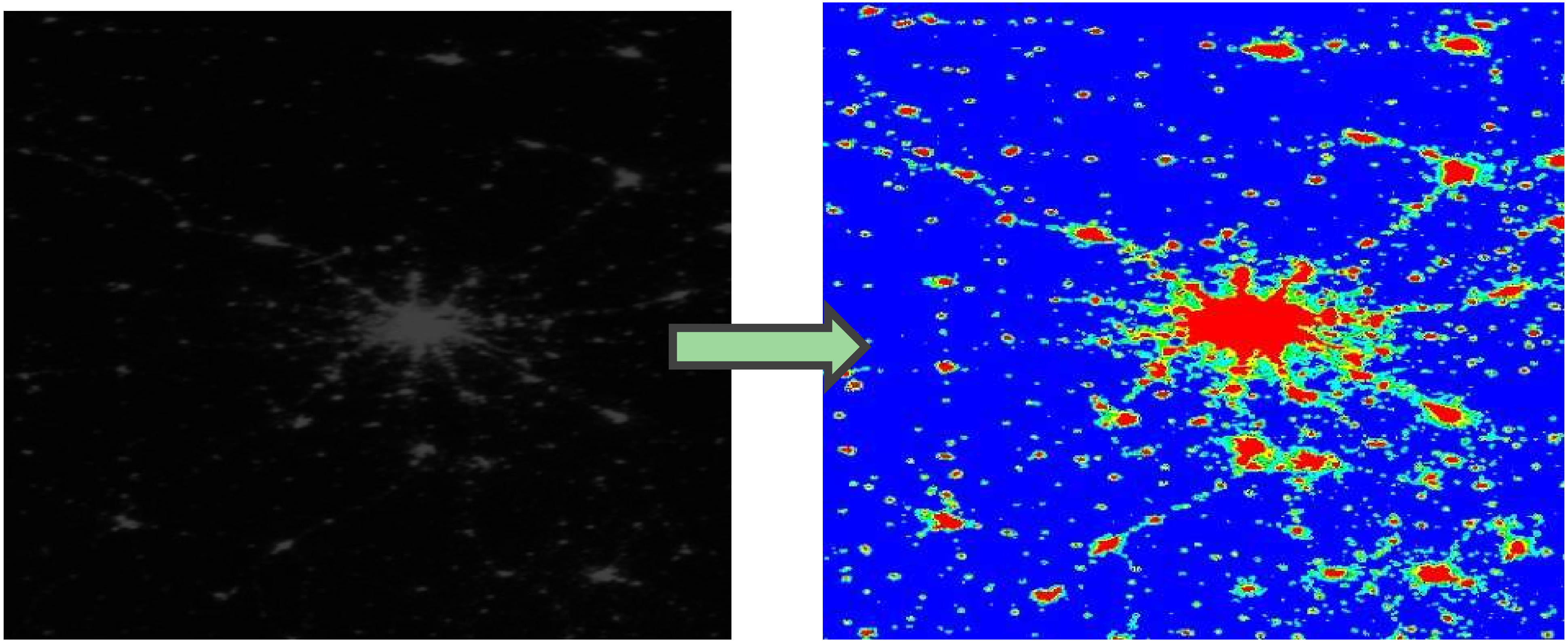
2. to predict the future spread of light

NTL MAP AND POPULATION DENSITY MAP LOOK THE SAME

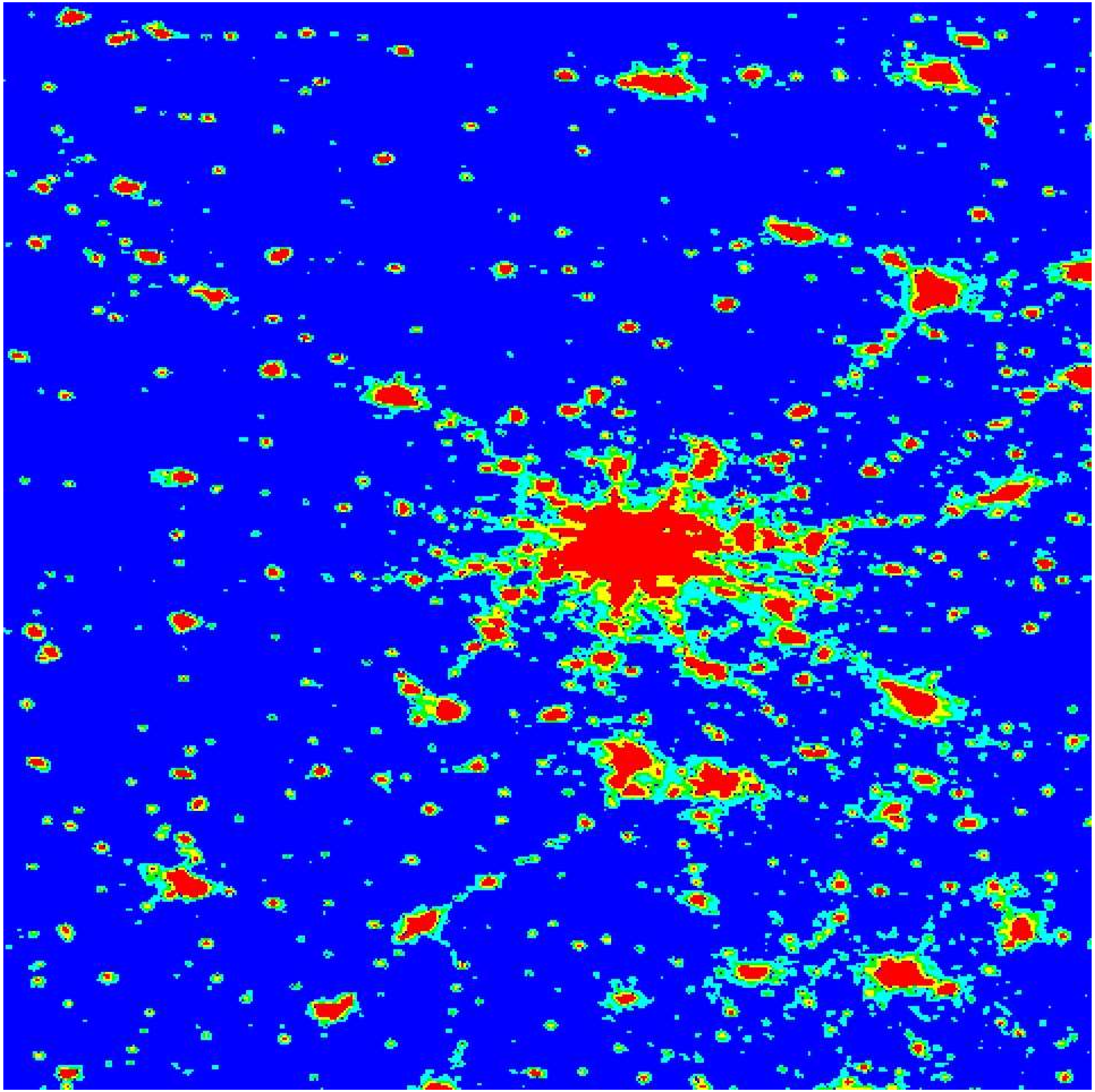


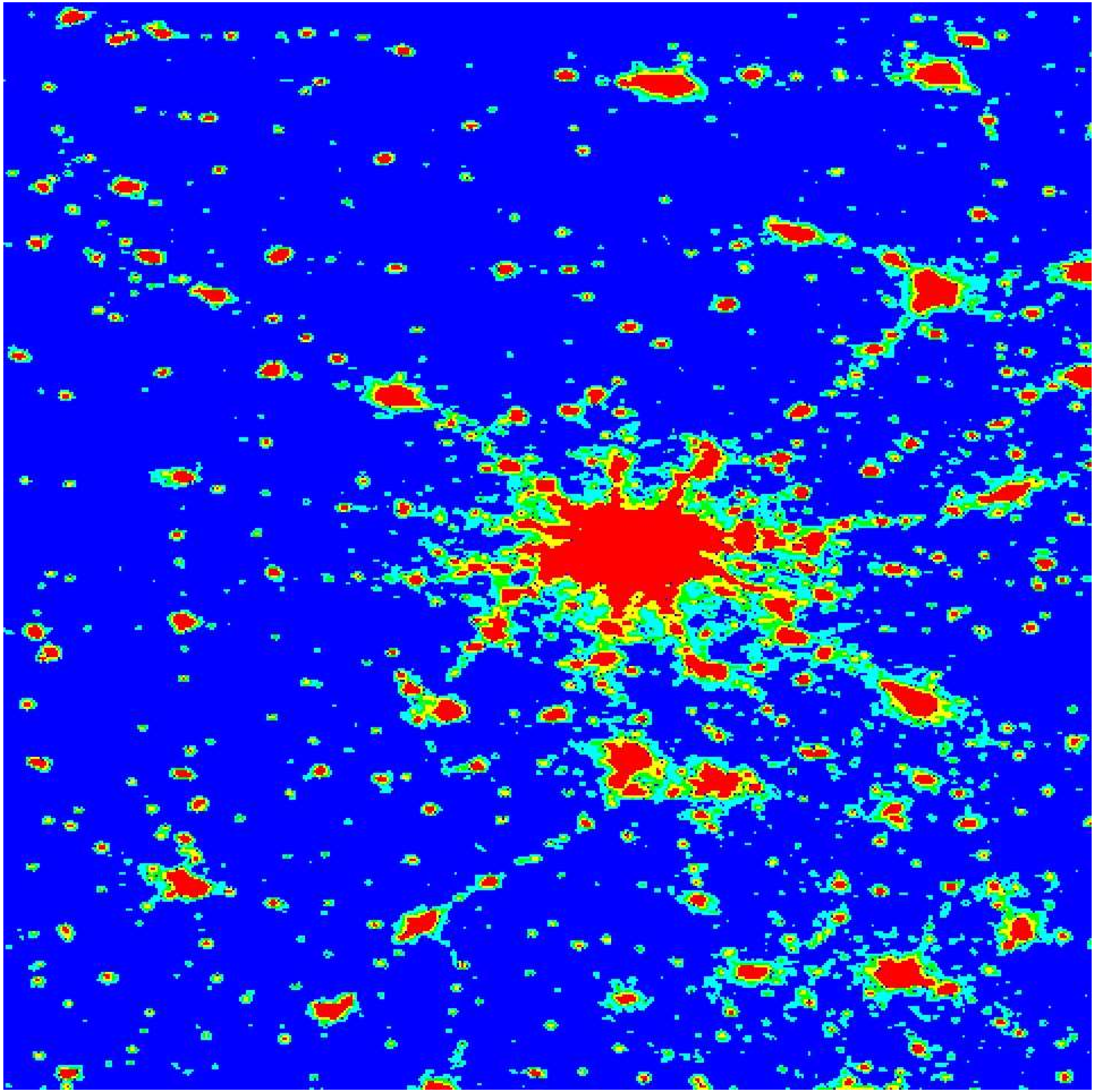
Creating maps of urbanization

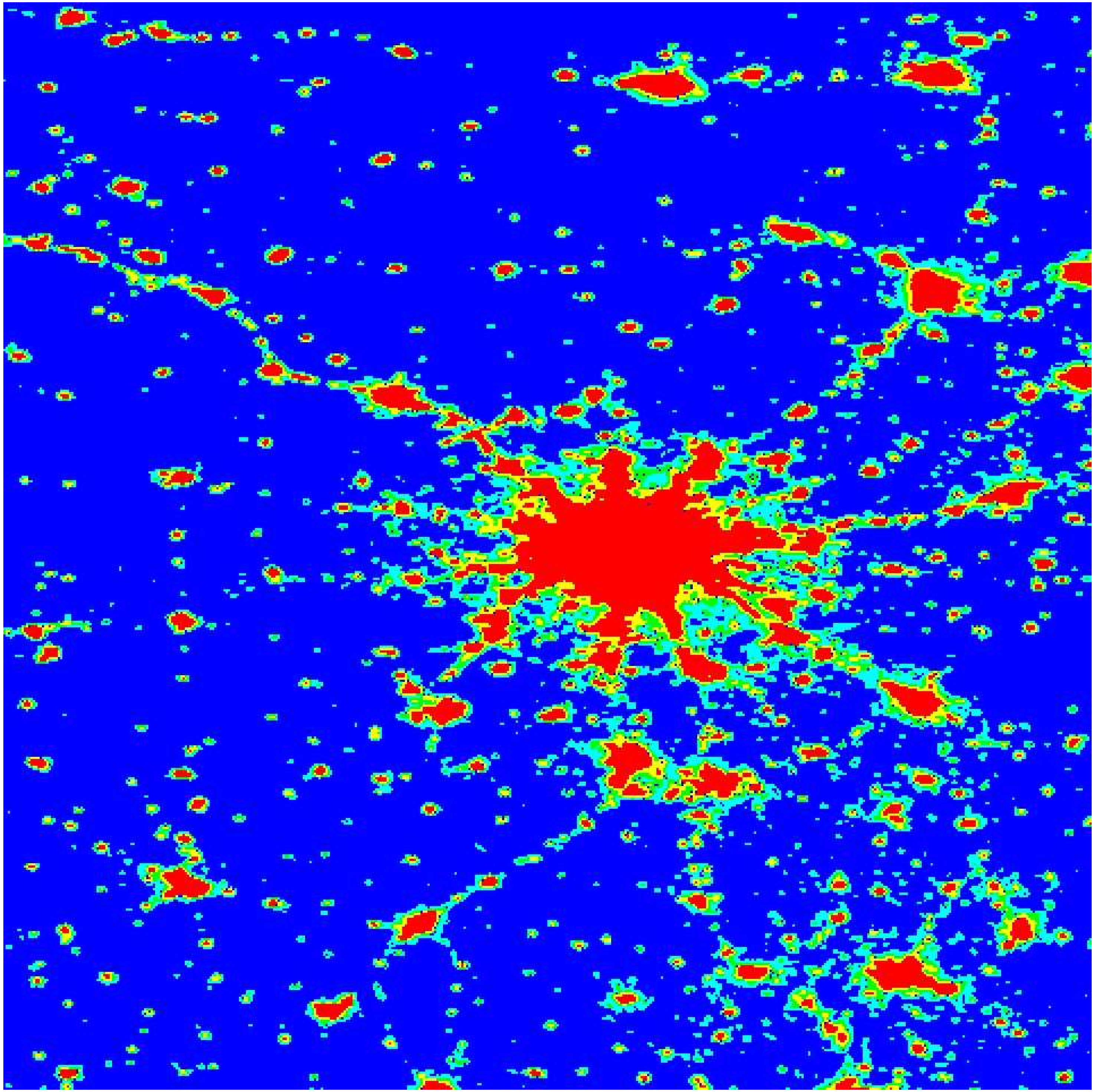
We divide the light into “five shades of gray” and use pseudo colors for visualization.
Colors show “levels of urbanization” (from the city center to outskirts):
(1) red -> (2) yellow -> (3) green -> (4) light blue -> (5) dark blue

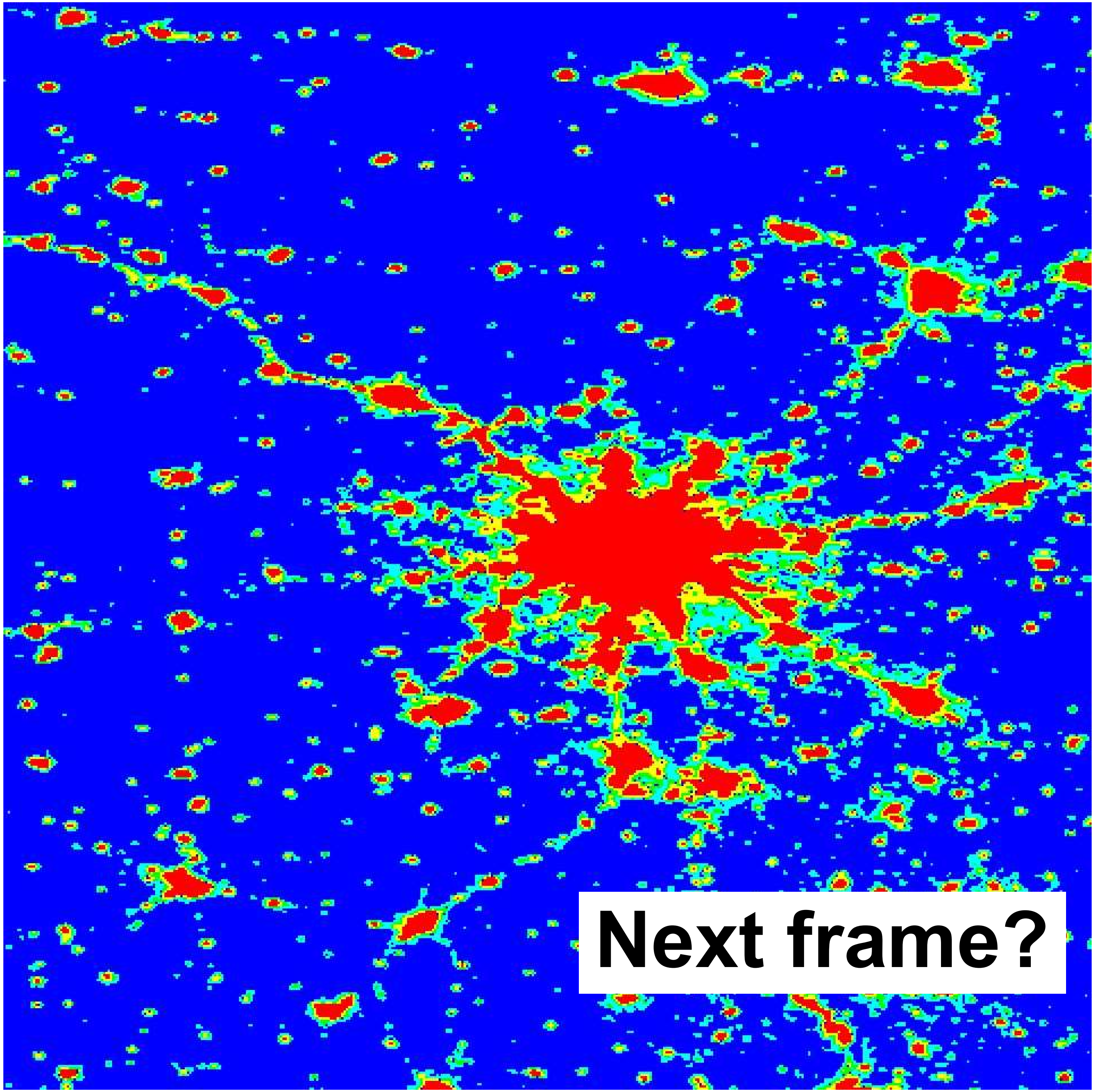


Moscow region (in 2008).








Next frame?

What will be the next frame?

To guess the script of a movie, you better need to know the laws of the genre. However, nowadays machine learning might help greatly.

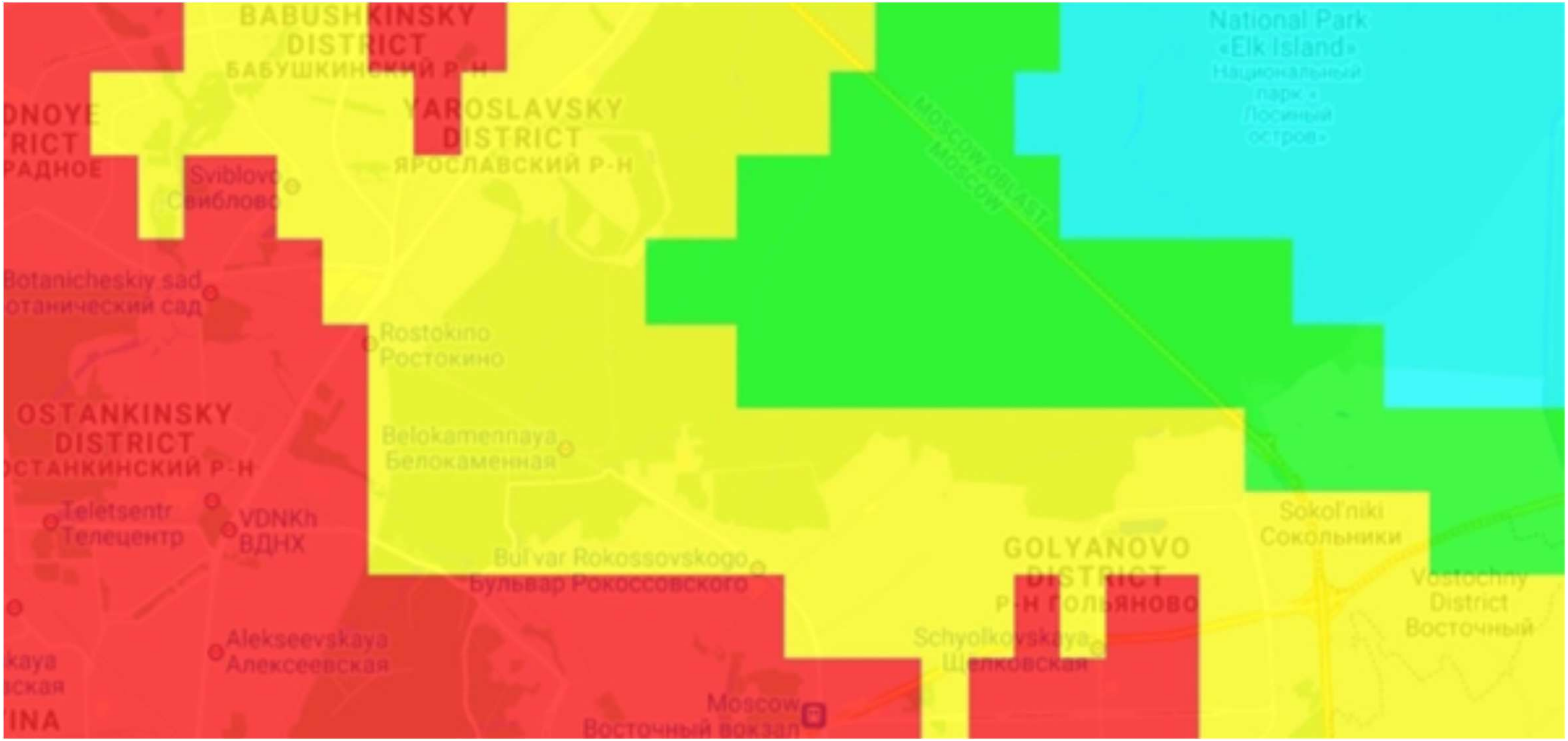
Explicit knowledge about the genre of the movie embedded in our model allows us to quickly start its functioning. Our software allows relatively fast production of movies for cities with similar economic and geographical conditions.

Using Machine Learning

- allows you to optimize the parameters of the model and in the future make it less dependent on explicit knowledge,
- allows you to evaluate the quality of modeling,

URBAN SPRAWL MODELLING

Computational scheme of evolution – probabilistic cellular automaton
The figure shows a fragment of the "urbanization map" (based on the brightness map of "night lights") in the north-east of Moscow in the Losiny Ostrov Park area. Different "levels of urbanization" are highlighted in color from the city center (red) to the outskirts (green and blue in most of the park).



Our model is a probabilistic cellular automaton with local data

- After being divided into cells, the time series of urbanization maps can be interpreted as states of a cellular automaton on a two-dimensional integer lattice.

- The difference between our cellular automaton and the Conway's Game of Life
  - In our model, cells can have five different states.
  - The transition rules are not deterministic, but probabilistic. To obtain a forecast, the Monte Carlo method is also used: numerous runs of the model of the evolution of the city from the initial state to the desired point in time through intermediate states. The final forecast is calculated based on the "averaging" of the set of calculated scenarios of the evolution of the city; this forecast is more stable than the forecasts obtained by a single operation of the cellular automaton
  - The "field of action" of the city evolution model is not a "homogeneous" two-dimensional integer lattice, but is a model of the earth's surface, and, accordingly, the rules for the transition of a cell to a new state may vary depending on the "place".

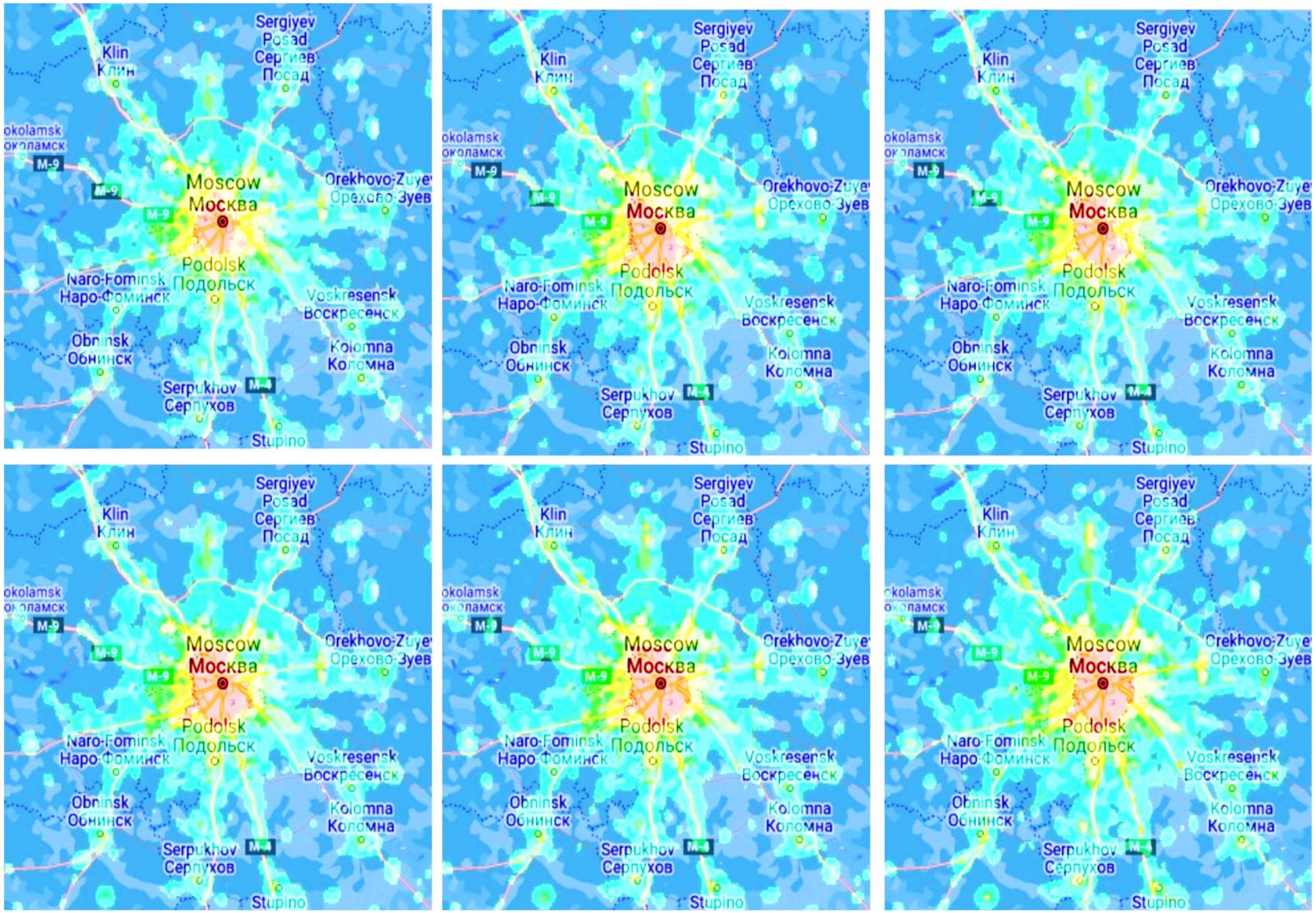
A TEST EXAMPLE OF A FORECAST

Training data set

2004, 2005, 2006

Computed forecast

2007, 2008, 2009

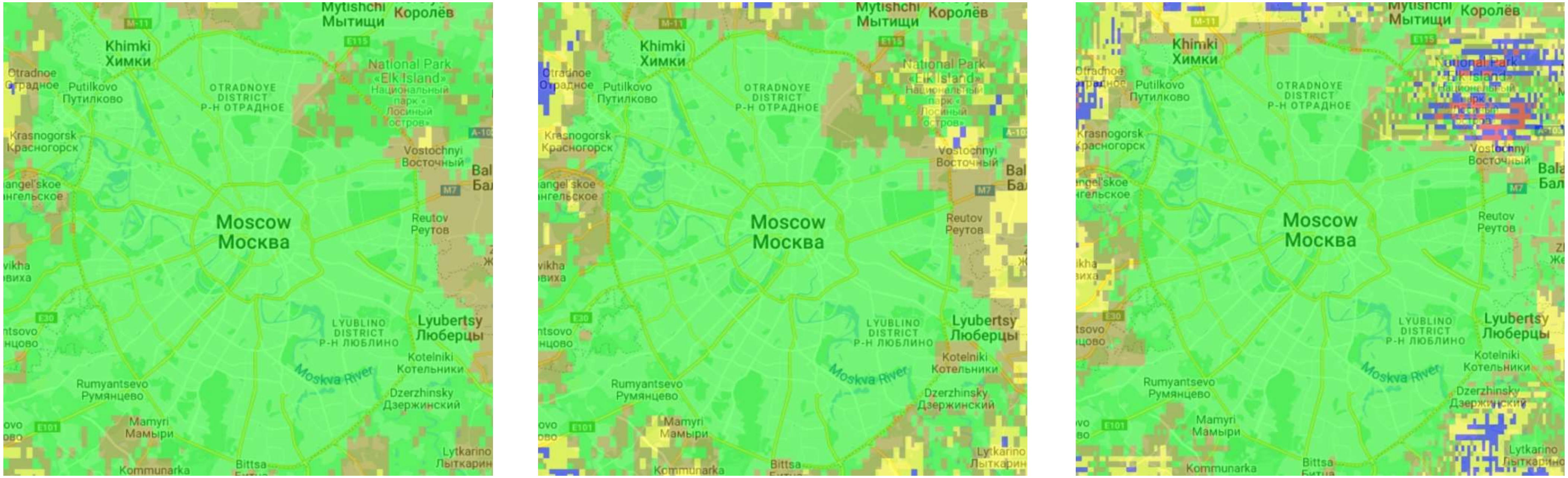


Quality of the forecast is estimated. We take our original dataset, split it into two parts. Train the algorithm on the first part, make predictions on the second part and evaluate the predictions against the expected (actual) results.

ASSESSMENT OF FORECAST QUALITY

A comparison of the forecast made, and the actual development of the city gives a quantitative assessment. Visualization of discrepancies allows you to understand the causes of discrepancies and ways to improve the model.

A TEST EXAMPLE OF THE FORECAST QUALITY ASSESSMENT



The training was done on the data of 2007, 2008, 2009. The forecast was calculated for the next three years. The figure shows the discrepancy between the forecast and the real development from left to right: 2010, 2011, 2012.

The colors on the maps show the level of discrepancies: Green - error < 2%, Brown < 5%, Yellow < 10%, Blue < 20%, Red > 20%

The Use Of Daytime ERS Data

The quality of forecasting improves when additional terrain and land use data are included in the model

Moscow and the surrounding area 2003

The figure shows the moisture content index (Normalized Difference Water Index, NDWI), calculated by us from the data of the Landsat channels.

1

SUMMARY

- Our group has developed software for a full cycle of urban sprawl modeling - from raw data processing to building a forecast of city growth and evaluating the quality of this forecast. As well as auxiliary tools for analyzing and visualizing data and results.
  - The prototype works fast and shows encouraging results in the quality of modelling.
  - The main direction of quality improvement at the moment is the introduction of additional data into the model.

- Our article in the proceedings of the conference contains mathematics on the use of machine learning in our model

SUMMARY (cont.)

- Our approach to urban growth prediction:

  – The problem considered is the prediction of the future growth of cities using probabilistic cellular automata (PCA) and historical satellite VIIRS data.

  – The approach is based on finding the "optimal values" of parameters for PCA model for accurate prediction of future urban growth. The problem is reduced to some machine learning optimization problem with an appropriate loss function: the difference between the predicted urban growth and actual urban growth on historical data.

    An appropriate regularization loss function makes the model more "stable" and robust that penalizes the complexity of the model, and generally helps to prevent overfitting.

  – To solve the optimization problem efficiently, it is possible to use the stochastic gradient descent method (SGD) or one of its variants.

Alexander Troussov, Ph.D.,
E-mail: troussov@gmail.com

Thank you!