

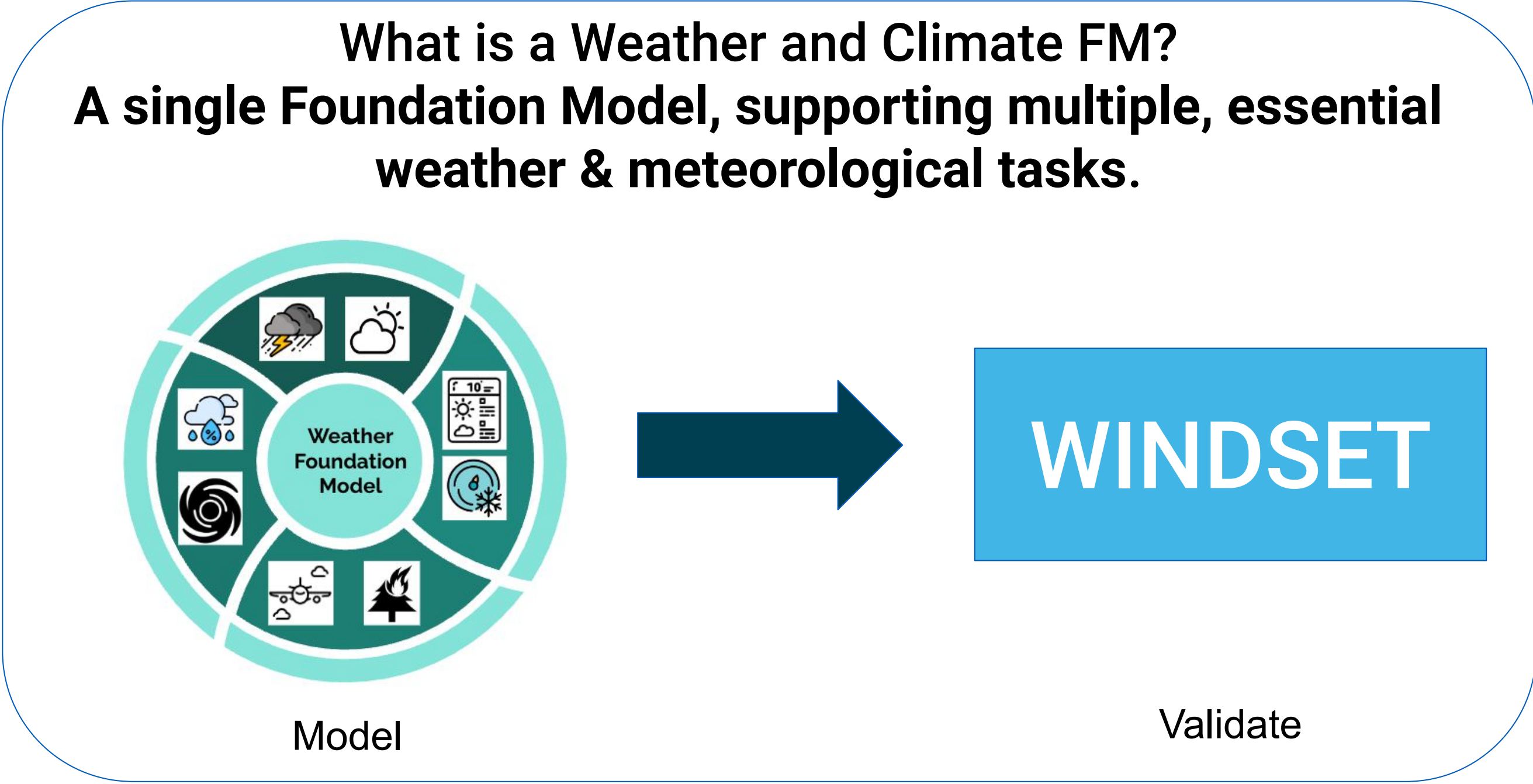
Rajat Shinde¹, Christopher Phillips¹, Sujit Roy^{1,2}, Aman Gupta³, Aditi Sheshadri³, Manil Maskey², Rahul Ramachandran²

sujit.roy@nasa.gov

1 - Earth System Science Center, The University of Alabama in Huntsville, 2 - NASA Marshall Space Flight Center, 3 - Doerr School of Sustainability, Stanford University, Stanford, CA, USA

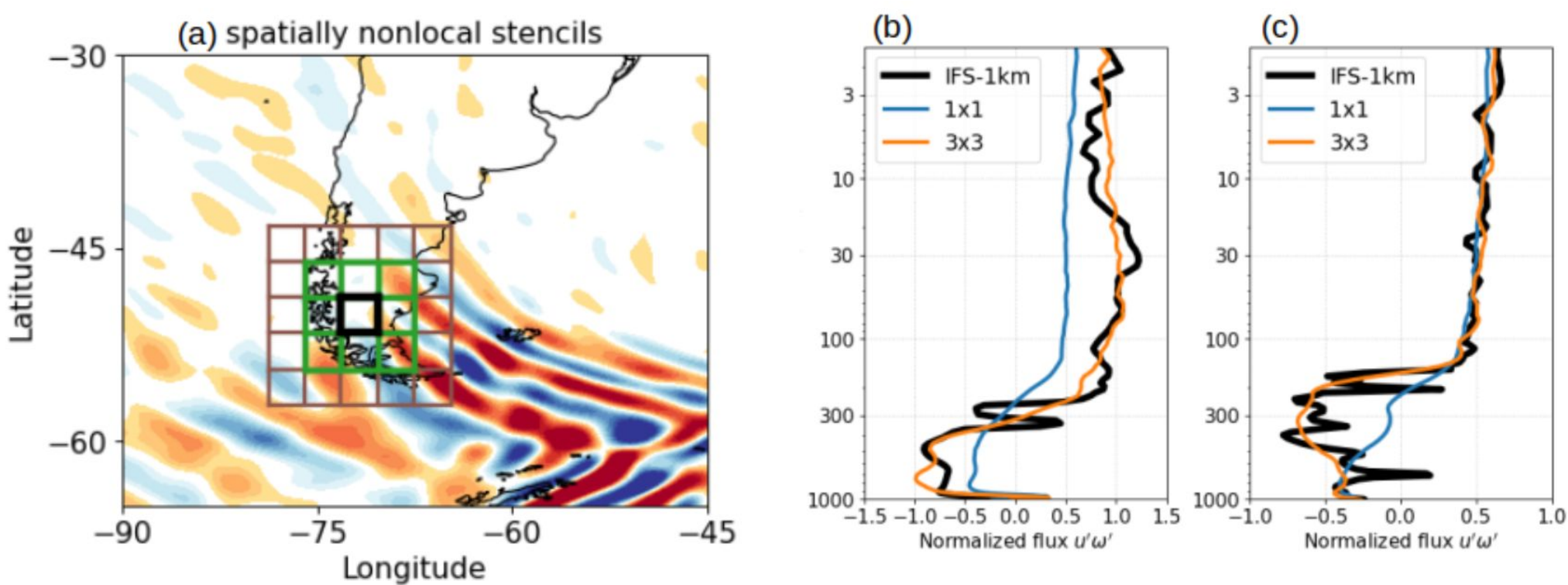
Overview

1	What?	Suite of Multi-resolution multi-modal datasets for fine-tuning Weather Foundation Model
2	Why?	Evaluating Efficacy of Weather FM across variety of tasks to explore generalizability
3	Need	Lack of specific, task-oriented AI-ready datasets for validating FMs



Datasets Details and Technical Validation

1. Non-Local Gravity Wave Parameterization (4-D Array)



2. Natural Language Weather Forecast Reports Generation

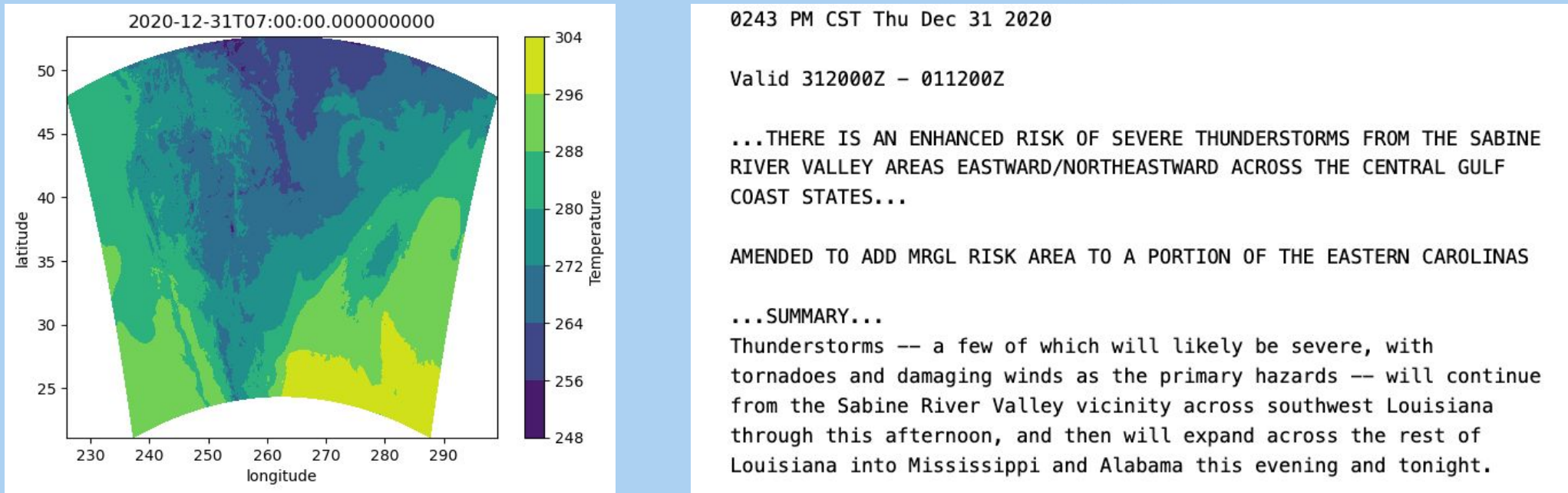


Fig. Illustration of training data for (a) temperature (b) corresponding forecast report for December 31, 2020

Input: Forecast discussion as text
Output: A model that can generate forecast discussions based on Weather FM generated reanalysis products
Source of the labels: SPC and NWS weather discussions
Number of Samples: > 7300

Temporal Extent: 2003-Present
Spatial Extent: CONUS for SPC; Regional for NWS

3. Aviation Turbulence Detection

Airports Overlayed on Density Map of PIREPs with Turbulence 2010-2020

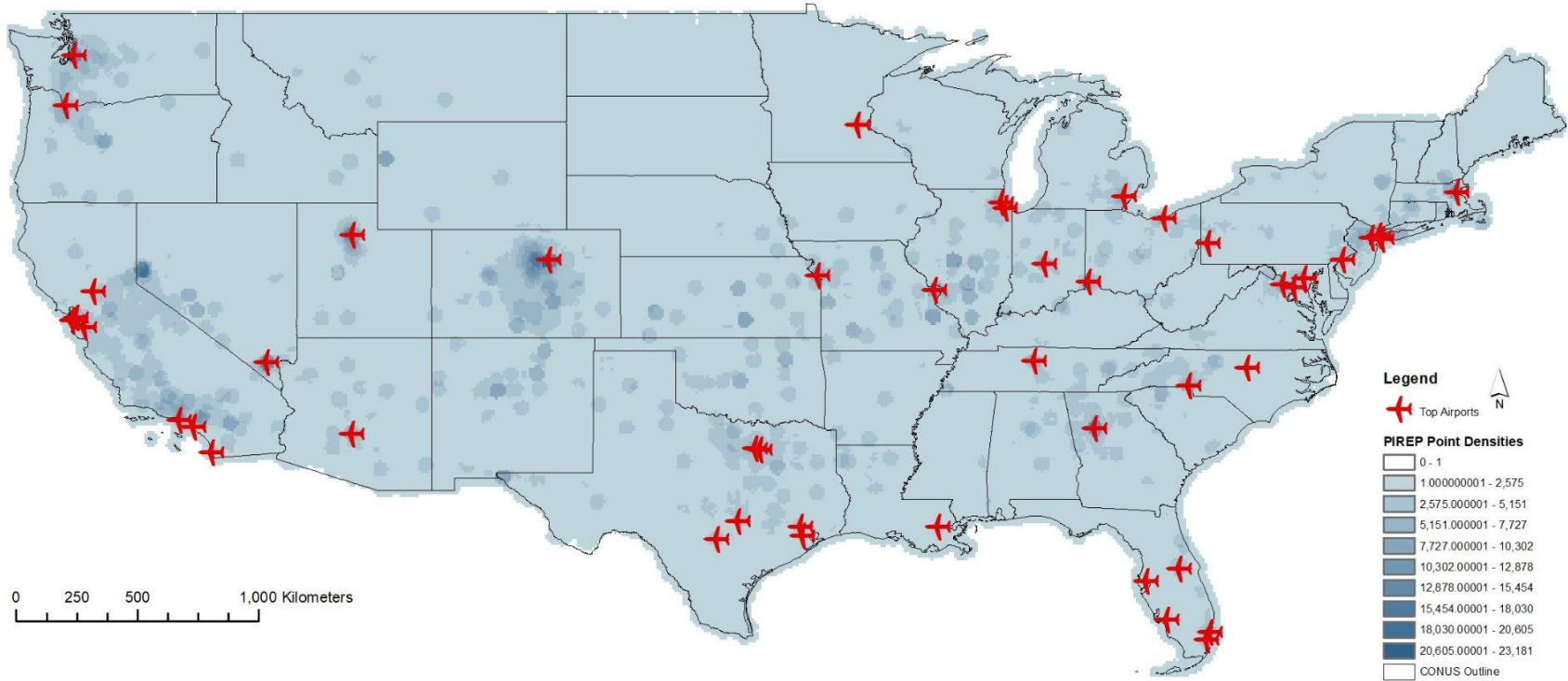


Fig. Density map of turbulent pilot reports overlaid with the 47 major airports over the contiguous U.S.

Input: Single MERRA 2 time step
Output: A model to predict turbulence risk and likelihood
Source of the labels: PIREPs archive provided by Iowa State University
Number of Samples: > 6,000,000

Temporal Extent: 2003-Present
Spatial Extent: United States

4. Searching Weather Parameter Analogs

