# RMS® CCRA® Training Program
## Geocoding & Hazard Retrieval Exercise: Northeast U.S. Windstorm Portfolio

**ANSWER KEY**

**Learning Objectives:**

This exercise demonstrates the impact that address accuracy and geocoding resolution have on exposure data analysis applications.

The following files are available to you to answer the subsequent questions:

- GeoHaz Exercise Data.xls: Exposure data from a RiskLink©-DLM EDM. Includes data from EDM address table, EDM property table, EDM loccvg table, EDM hudet table, and RDM locstats table.
- NE_WIND_VALUExD2C.pdf: Distance to coast summary report
- GeoStan_StatusCodes v19.00.pdf: Centrus GeoStan geocoding status codes output into the RiskLink EDM to record address changes during the geocoding process and spatial accuracy.

**PART I – Exposure Data Address Accuracy and Geocoding Resolution Review**

You have been presented with a RiskLink-DLM EDM that has a portfolio containing 98 locations throughout the Northeast United States. It has 17 accounts and it includes a variety of commercial and residential properties. Prior to geocoding the portfolio, you review the address information in the address table of the EDM and note that there are a few questionable addresses shown in the table below.

| LOCNUM | Address/Description | LOCNUM | Address/Description |
|---|---|---|---|
| 24 | New Heaven should be New Haven | 1039 | 182222 is an unlikely street number |
| 99 | Universalblvd should be two words | 985 | No ZIP Code |
| 332 | Blk Horse Pike should be Black Horse Pike | 1284 | ZIP Code 67640 is from out-of-state |
| 56 | Nework should be Newark | 272 | East Lime should be East Lyme |

Some geocoders, including those used in RiskLink, will automatically correct some address errors, so you geocode the portfolio in RiskLink. The resulting address table with the geocoding results is shown on the Geocoded Data sheet of the GeoHaz Exercise Data.xls spreadsheet.

1. What geocoding resolution(s) did you achieve for the locations in this portfolio?

   *LOCNUMs 99 and 1039 returned a GeoResolutionCode of 5, which indicates a postal code level match. All 96 of the other locations returned match levels of 2, indicating high-resolution matches (includes street, parcel, and building level matches). In addition, eight locations also returned Sanborn building IDs, signifying building-level matches.*

2. Based on the geocoding level results, which of the previously observed errors were corrected? List the locations which were not corrected and provide a probable explanation as to why.

   *All were automatically corrected by the geocoding software, with the exception of LOCNUMs 99 and 1039. The software was not able to correct Univeralblvd, because it looked for this spelling and there was nothing that was similar enough for it to be able to make the correction during the standardization process. The software was also not able to correct 182222 because it could not tell whether the user had entered this as a valid street number.*

3. Unit 1 covers spatial uncertainty in address geocoding, and its influence on exposure data analysis and modeling results. Ideally, you want to send your locations into the catastrophe models having eliminated as much uncertainty in geocoding as you have control over. That said, you have achieved street level geocoding in almost all of your locations. Would you recommend fixing the remaining problems to improve these geocoding results before you perform the modeling process on this portfolio, and if so, how?

   *A high-resolution geocoding "hit rate" of 98% by location (96 locations out of a total of 98) and 94% by value ($1.17B out of $1.25B) is exceptionally good. That being said, it is important to remember that any information gleaned from a postal code level geocode rests on a number of assumptions. The first is that the centroid of the postal code – weighted or not – is representative of all locations in that postal code. The second is that hazard values (in this case distance to coast, elevation, surface roughness, etc.) are uniform across the postal code.*

   *These assumptions may be reasonable within certain circumstances. But where important decisions are determined by any of these assumed values, it may be advisable to pursue a higher-level address match if at all possible.*

**PART II – Assessing Geocoding and Hazard Uncertainty through Output Codes**

Geocoding process changes are recorded in the EDM for some regions and geocoding levels. The next part of the exercise will use the data in the GeoHaz Exercise Data.xls spreadsheet and the GeoStan_StatusCodes v19.00.pdf document to assess geocoding uncertainty. The following data fields exist in the address table of the EDM:

**GeoMatchCode** – GeoStan match codes that indicate the portions of the address that matched or did not match to the GeoStan directory file. The code digits do not specifically refer to which address elements did not match, but rather to why the address did not match.

**GeoLocationCode** – This is the GeoStan location code and is an indicator of spatial accuracy of the assigned geocode.

**GeoBuffer** – Uncertainty code for a location geocoding to a high-resolution (in feet).

**GeoResolutionCode** – RMS geocoding match level (2 = high resolution [street, parcel or building]; 5 = postal code level)

4. Review the data in the GeoHaz Exercise Data.xls spreadsheet. Is there a relationship between the GeoMatchCode, GeoLocationCode, GeoBuffer, and GeoResolutionCode? If so, what is the relationship? Provide an explanation.

   *The GeoBuffer values are only filled for locations that geocoded to the street, parcel, or building levels (GeoResolutionCode = 2). This is a high-resolution code, and therefore any locations with a lower geocoding resolution (e.g., GeoResolutionCode = 5) do not contain information in the GeoBuffer field. In addition, locations that did not geocode to the street, parcel, or building level have a GeoMatchCode beginning with the letter E and a GeoLocationCode beginning with the letter Z, both of which indicate that no street address match was found.*

5. What do you notice about the relationship between the GeoMatchCode, GeoLocationCode, and GeoBuffer codes in the GeoHaz Exercise Data.xls spreadsheet?

   *The GeoBuffer values are never less than 50 feet and the highest value is nearly 540 feet. The table following shows the combinations of GeoMatchCode and GeoLocationCode values for this dataset. You will notice that the building level geocode receives the minimum GeoBuffer value of 50 feet. GeoMatchCode values that indicate changes in the city and/or postal code may still provide a very low GeoBuffer value. This indicates that the accuracy of the geocoding resolution may be suspect while the calculated spatial uncertainty is low. Finally, note that the street intersection address has a lower GeoBuffer value than many street address geocodes. In this case, the interpolated geocode for long street blocks carries more uncertainty than the geocoded location for a street intersection.*

*GeoMatchCode, GeoLocationCode values:*

| | |
|---|---|
| *A8A, AS0* | *A8A = Address matches to alias record. ZIP+4 changed. Street name and pre-directional changed.*<br>*AS0 = Location on a street range. House range address geocode. Most accurate geocode available for street range.* |
| *S80, AP02* | *S80 = Match found in USPS data. ZIP+4 changed. No change in address line.*<br>*AP02 = Point-level data location. Parcel centroid.* |
| *S80, AP05* | *S80 = Match found in USPS data. ZIP+4 changed. No change in address line.*<br>*AP05 = Point-level data location. Center of an addressable building footprint polygon.* |
| *S80, AS0* | *S80 = Match found in USPS data. ZIP+4 changed. No change in address line.*<br>*AS0 = Location on a street range. House range address geocode. Most accurate geocode available for street range.* |
| *S80, AS2* | *S80 = Match found in USPS data. ZIP+4 changed. No change in address line.*<br>*AS2= Location on a street range. House range address geocode. Changed original segment name to match the USPS spelling.* |
| *S80, AS3* | *S80 = Match found in USPS data. ZIP+4 changed. No change in address line.*<br>*AS3 = Location on a street range. House range address geocode. Changed original segment name to match the USPS spelling and interpolated in a street segment that did not initially contain address ranges.* |
| *S89, AS0* | *S89 = Match found in USPS data. ZIP+4 changed. Street name and street type changed.*<br>*AS0 = Location on a street range. House range address geocode. Most accurate geocode available for street range.* |
| *SA0, AP02* | *SA0 = Match found in USPS data. City and ZIP+4 changed. No change in address line.*<br>*AP02 = Point-level data location. Parcel centroid.* |
| *SA0, AS0* | *SA0 = Match found in USPS data. City and ZIP+4 changed. No change in address line.*<br>*AS0 = Location on a street range. House range address geocode. Most accurate geocode available for street range.* |
| *X000, AX8* | *X000 = Match found is an intersection of two streets. No change in address line.*<br>*AX8 = Intersection geocode. Interpolated (divided-road) intersection of geocode.* |
| *E020, ZC5X* | *E020 = Error or no match. No matching streets found in directory.*<br>*ZC5X = Location derived from a ZIP centroid. Unclassified Cenus accuracy. ZIP Code centroid. Less than 80% of addresses in this ZIP Code are in a single Census Tract – assigned to ZIP Code centroid.* |
| *E022, ZC5X* | *E022 = Error or no match. No matching segments.*<br>*ZC5X = Location derived from a ZIP centroid. Unclassified Cenus accuracy. ZIP Code centroid. Less than 80% of addresses in this ZIP Code are in a single Census Tract – assigned to ZIP Code centroid.* |

6. Compare the DISTCOAST values with their associated DSCTMATCH values and GeoResolutionCode values in the GeoHaz Exercise Data.xls spreadsheet.

   **DISTCOAST** – Distance to coast in miles

   **DSCSTMATCH** – Hazard match resolution

   How do the distance to coast resolutions (DISTCOAST), hazard match resolutions (DSCSTMATCH), and geocoding match (GeoResolutionCode) appear to relate to each other? Provide an explanation.

   *Locations that geocoded at the street level got DSCSTMATCH = 2, unless they were more than 10 miles from the coast. Street level geocodes that were more than 10 miles from the coast got DSCSTMATCH of 143-144, which represent VRG cells of various sizes. Locations that geocoded at the postal code level got DSCSTMATCH = 5. However, even if they were more than 10 miles from coast, they would have gotten the same DSCSTMATCH because a postal code geocode cannot receive a higher level of hazard match.*

**PART III – Geocoding Impact on Portfolio Conformance to Underwriting Guidelines**

Your company-wide underwriting guidelines aim to restrict coastal business throughout the United States. These guidelines establish maximum limits for bands of territory starting at the coast and moving inland. In particular, the guidelines limit coastal exposure in the Northeast to the following limit amounts:

**Maximum Coastal Limits by Distance to Coast Band for the Northeast U.S.**

| Distance to Coast Bands* | Maximum Limits ($M) |
|---|---|
| 0.0 – 0.3 miles | $15,000,000 |
| 0.3 – 1.0 miles | $75,000,000 |
| 1.0 – 5.0 miles | $200,000,000 |
| 5.0 – 10.0 miles | $500,000,000 |
| 10.0+ | unlimited for wind |

**Note: The top range is NOT inclusive of the maximum value. In other words, the 0.0 – 0.3 mile band considers all locations that have distance to coast matches from 0.0 miles up through 0.2999 miles. Those with distance to coast of 0.3 miles will fall into the second band.*

Your job is to figure out how to make your portfolio conform to these guidelines – if possible. To identify the current exposure, you refer to a recently run distance to coast summary report (NE_WIND_VALUExD2C.pdf). You also have the relevant exposure and results information available to you to as you start to determine if the portfolio is within the guidelines. (Note: Relevant information from the EDM and RDM is provided in the GeoHaz Exercise Data.xls spreadsheet)

7. Does your northeast portfolio comply with the new guidelines? If not, which bands are over the maximum allowed limits?

   *The portfolio does not comply with the new guidelines because the 0.3-1.0 mile and the 1.0-5.0 miles distance to coast bands have significantly more exposure than the allowed amount of exposure. All other bands are within their limits.*

8. You need to present this information to your Chief Underwriting Officer this afternoon. What additional information (if any) will you include with the exposure summary to help him understand the exposure in this region? Think about whether all geocoding uncertainty has been eliminated in your results, and what confidence level you have in the quality of the location data and geocoding results you have used in your analysis.

   *The main concern to highlight to your CUO is the lack of precise spatial information for the two addresses that have geocoded only to the postal code level. These represent only a "guess", particularly since they are in zones relatively near the coast. This may not be of grave concern in inland areas, where wind hazard is relatively homogeneous, or for locations that are relatively low in value; but near the coast, where decisions of inclusion and non-inclusion hang in the balance, it may be unwise to make determinations based, even in part, on low-resolution geocoding results.*

   *Since a change in geocoding level may also alter our perception of the way in which our exposure is distributed, we should recommend that, at a minimum, we capture more detail on these two locations. Furthermore, since important decisions are to be made based on distance to coast data, we may recommend to our CUO that we update our underwriting practices to require high-resolution address information in specific coastal zones, or in other areas where wind hazard is highly variable.*

After reviewing the report, your CUO asks you to make recommendations about how to make this region comply with the new guidelines. Before you make your recommendation, you have decided to review the portfolio at a closer level to make sure that you are not making a decision based on inaccurate information. You decide that the greatest concern in using the exposure information is that there are two locations with postal code resolution geocoding.

9. List two reasons why using these locations with postal code geocoding resolution to make a critical company decision regarding your portfolio is of concern to you. Look not only at the quality of the address data but also at the geocoding results and the limits of each exposure.

   *Of greatest concern is that the coastal exposure guidelines do not account for data quality or geocoding match level. In other words, some policies are being set up for non-inclusion based on their locations' distances to the coast, without accounting for what those values actually represent. The second concern is related: that a location worth more than $77 million lies in the 0.3-1.0 mile distance to coast band, but has geocoded only to the postal code level.*

In order to raise your confidence level in the data, you decide to improve the geocoding match for the postal code level locations and then re-run an AAL (average annual loss) analysis to observe whether an improved match would affect the overall risk profile for the portfolio. You correct the addresses that appeared incorrect and that resulted in postal code geocoding resolution matches, namely the "Universalblvd" of LOCNUM 99 and the "1822222" of LOCNUM 1039. The first is easy to correct – i.e., to change it to "Universal Blvd"; but the second may require more research. After some research, you find that "182 Calcutta St" is the valid address. Both corrections now yield high-resolution geocoding matches.

The distribution of limits by distance to coast band is shown in the table below for the original geocoded portfolio as well as the updated portfolio (i.e., the re-geocoded portfolio after correcting LOCNUMs 99 and 1039).

| Distance to Coast Band | Maximum Limits ($M) | Original Limits ($M) | Updated Limits ($M) |
|---|---|---|---|
| 0.0 – 0.3 miles | $15,000,000 | $13,992,248 | $13,992,248 |
| 0.3 – 1.0 miles | $75,000,000 | $159,435,653 | $81,705,653 |
| 1.0 – 5.0 miles | $200,000,000 | $451,197,039 | $528,675,919 |
| 5.0 – 10.0 miles | $500,000,000 | $190,183,792 | $190,183,792 |
| 10.0+ | unlimited for wind | unlimited for wind | unlimited for wind |

10. Explain why the original limits and updated limits columns have different values.

    *Re-geocoding the two locations previously at the postal code level yields a very different distance to coast profile. Now the 0.3-1.0 mile band, which once had more than double the allowed insured values, is only slightly above its allowed limit. The change in geocoding now puts the 1.0-5.0 mile band over its allowed limit by nearly $328 million. That said, the overall distribution of value has "moved" inland.*

11. Taking any other factors into account (for example, location level hazard or loss values within the prescribed distance to coast bands), make some side notes for your final report. Based on what you have learned today, would you recommend that your company continue with windstorm underwriting guidelines based solely on distance to coast? Justify your answer with a description.

    *Ultimately, we have three choices regarding the distance to coast bands where our exposure exceeds the guidelines: accept the risk, non-renew the risk, or cede some or all of the risk to a reinsurer. Before we make any final decisions (the "correct" decision is ultimately a question of risk appetite), there is one additional component that should be investigated.*

    *Distance to coast is a good starting point for understanding windstorm risk; however, localized wind speeds are highly impacted by the hazard that exists between the origin of the wind and the property that we are evaluating. As a result, we may wish to analyze each property's exposure to the wind peril before deciding which locations, if any, require action. In addition, it is important to keep in mind that the RMS coastline has been developed to take the flood extents of the storm surge model into consideration and it encompasses water bodies and river channels where the width is greater than 50 meters.*

    *For example, based on the modeled ground up AAL in the table (taken from the locstats table in the RDM) compared to the distances to coast (from the hudet table in the EDM), we notice that some locations with very similar distances to coast have very different modeled losses. Even in the first and second distance to coast bands, for example, nearly identical site hazard values are presented side-by-side with very different loss costs. Consider the following:*

| LOCNUM | Dist. to Coast | Value | GU AAL | Loss Ratio (per mille) |
|---|---|---|---|---|
| 5 | 3.033 miles | $57.0M | $1,677 | 0.029 |
| 12 | 2.733 miles | $77.2M | $18,170 | 0.235 |

    *LOCNUM 5 is located in Montgomery County, Pennsylvania, near Philadelphia – just across the Delaware River. LOCNUM 12 is located in Plymouth County, Massachusetts – just south of*

*Boston. The first is off of a well-protected river near an estuary, some 65 miles from the Atlantic Ocean, while the second is adjacent to the open Atlantic Ocean.*

*So, while distance to coast provides us with a decent starting point in terms of our exposure to the peril of wind, we must note that any two risks that fall within the same 'band' may have significantly different exposure to the actual peril. This additional knowledge can either be used on a reactive basis (e.g., which policies should we non-renew or purchase reinsurance for) or could be included on a proactive basis (e.g., model each risk and require a maximum loss cost threshold by region).*

***Side Note:** If you were to review tables in the RDM, you would find the maximum single-event loss for the Massachusetts location (LOCNUM 12) to be $6.5M, which is a loss cost of $84.46 per mile, while for the Pennsylvania location (LOCNUM 5) it is $2.9M, which is a loss cost of $50.77 per mile. In fact, there are ten events that cause losses to the Massachusetts location that are in excess of the maximum single-event loss for the Pennsylvania location – all this despite the fact that the Massachusetts and Pennsylvania locations are very similar distances inland using the distance to coast measurement.*