Storms across the USA: their impact in population health and economic consequences after the ‘National Weather Service Instruction 10-1605’ from August 17, 2007, NOAA

Luis David Bedon Gomez

9 Mai 2017

Synopsis

The following report discuss the significance of storms concerning population health and economic impact across the USA, based on storm-data sets made public by the NOAA under the ‘National Weather Service Instruction 10-1605’ from August 17, 2007, that gives information collected from 1950 to 2007.

From the analysis can be concluded, that the most severe storm events affecting population health are the tornadoes, representing by a factor of 15 the most harmful type of storm, regarding both injuries and fatalities. After the tornadoes, a group of disasters like excessive heat, thunderstorm winds, floods and lighting are the most dangerous disasters.

The major economic impact comes from floods. According to the present NOAA data set, floods are responsible for an estimate of nearly 0,15 trillion USD on economic damages from 1950 to 2007. After floods, hurricanes, tornadoes and storm surges are responsible for economic losses between 400 to 700 billion USD each.

1. Data Processing

1.1 Downloading and loading the data

The data set is stored in a compressed csv-File “repdata%2Fdata%2FStormData.csv.bz2” that is loaded through the following routine:

```
setwd("~/DataScienceSpecialization/05ReproducibleResearch/RepData_PeerAssessment2")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)


if(!file.exists("repdata%2Fdata%2FStormData.csv")){
  urldata<-"https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file(urldata,destfile = "StormData.csv.bz2",method = "curl")
```

```r
  unzip("StormData.csv.bz2")
}

# NA-values are stored as "NA" if Rds-File does not exist
if(!file.exists("stormdata.Rds")){
  stormdata<-read.csv("repdata%2Fdata%2FStormData.csv", stringsAsFactors = FALSE, na.strings = "")
  # To avoid long reading times, save stormdata as Rds-File
  saveRDS(stormdata,file="stormdata.Rds")
} else stormdata<-readRDS("stormdata.Rds")
```

1.2 Preprocessing the data

Once the data set is loaded, the data are screened to understand the file using the following instruction:

```r
str(stormdata)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  NA NA NA NA ...
##  $ BGN_LOCATI: chr  NA NA NA NA ...
##  $ END_DATE  : chr  NA NA NA NA ...
##  $ END_TIME  : chr  NA NA NA NA ...
##  $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ COUNTYENDN: logi  NA NA NA NA NA NA ...
##  $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ END_AZI   : chr  NA NA NA NA ...
##  $ END_LOCATI: chr  NA NA NA NA ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  NA NA NA NA ...
##  $ WFO       : chr  NA NA NA NA ...
##  $ STATEOFFIC: chr  NA NA NA NA ...
##  $ ZONENAMES : chr  NA NA NA NA ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  NA NA NA NA ...
```

```
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

1.2.1 Extracting Data Concerning Public Health and Event Type

From the information about the file, its columns and its rows given by the "str"-function, the relevant data concerning the impact of storm events in public health has to be selected.

We see that the columns "Fatalities" and "Injuries" report the impact on public health. They belong already to the "numeric"-class of data.

The storm event information can be extracted from the column "Evtype" of class "character".

A new data set "pubh_evt" with the needed information is created using the following routine:

```
pubh_evt<-data.frame(EVTYPE=stormdata$EVTYPE,FATALITIES=stormdata$FATALITIES,INJURIES=stormdata$INJURIES
saveRDS(pubh_evt,file="pubh_evt.Rds")
```

1.2.2 Extracting Data Economic Damage and Event Type

Regarding the economic damage, the columns "PROPDMG/PROPDMGEXP" and "CROPDMG/CROPDMGEXP" are needed, which give the estimates of the damages on property and crops caused by the corresponding storm types.

As pointed on the documentation of the data set on page 12, section 2.7, the Property Damage "PROPDMG" and Crop Damage "CROPDMG" are notated as numbers that have to be multiplied by the factor contained in "PROPDMGEXP/CROPDMGEXP", corresponding to the exponent for thousands-"K", millions-"M" and billions-"B".

The following routine preprocess and extract the needed information:

```
# Reconvert PROPDMG and CROPDMG after the code in PROPDMGEXP and CROPDMGEXP.

stormdata$PROPDMG[grep("[K|k]",stormdata$PROPDMGEXP)]<-
  stormdata$PROPDMG[grep("[K|k]",stormdata$PROPDMGEXP)]*1e3
stormdata$PROPDMG[grep("[M|m]",stormdata$PROPDMGEXP)]<-
  stormdata$PROPDMG[grep("[M|m]",stormdata$PROPDMGEXP)]*1e6
stormdata$PROPDMG[grep("[B|b]",stormdata$PROPDMGEXP)]<-
  stormdata$PROPDMG[grep("[B|b]",stormdata$PROPDMGEXP)]*1e9

stormdata$CROPDMG[grep("[K|k]",stormdata$CROPDMGEXP)]<-
  stormdata$CROPDMG[grep("[K|k]",stormdata$CROPDMGEXP)]*1e3
stormdata$CROPDMG[grep("[M|m]",stormdata$CROPDMGEXP)]<-
  stormdata$CROPDMG[grep("[M|m]",stormdata$CROPDMGEXP)]*1e6
stormdata$CROPDMG[grep("[B|b]",stormdata$CROPDMGEXP)]<-
  stormdata$CROPDMG[grep("[B|b]",stormdata$CROPDMGEXP)]*1e9

econ_evt<-data.frame(EVTYPE=stormdata$EVTYPE,
                     PROPDMG=stormdata$PROPDMG,
                     CROPDMG=stormdata$CROPDMG)
saveRDS(econ_evt,file="econ_evt.Rds")
```

2. Data Analysis

2.1 Data Concerning Public Health and Event Type

To address this question, I use the following pipeline:

```r
pubh_evt_a<-pubh_evt %>%
  group_by(EVTYPE)%>%
  summarize(FATALITIES=sum(FATALITIES,na.rm=1),
            INJURIES=sum(INJURIES,na.rm=1),
            TOTAL=FATALITIES+INJURIES) %>%
  arrange(desc(TOTAL))
head(pubh_evt_a,15)
```

```
## # A tibble: 15 × 4
##               EVTYPE FATALITIES INJURIES TOTAL
##               <fctr>      <dbl>    <dbl> <dbl>
## 1            TORNADO       5633    91346 96979
## 2     EXCESSIVE HEAT       1903     6525  8428
## 3          TSTM WIND        504     6957  7461
## 4              FLOOD        470     6789  7259
## 5          LIGHTNING        816     5230  6046
## 6               HEAT        937     2100  3037
## 7        FLASH FLOOD        978     1777  2755
## 8          ICE STORM         89     1975  2064
## 9  THUNDERSTORM WIND        133     1488  1621
## 10      WINTER STORM        206     1321  1527
## 11         HIGH WIND        248     1137  1385
## 12              HAIL         15     1361  1376
## 13 HURRICANE/TYPHOON         64     1275  1339
## 14        HEAVY SNOW        127     1021  1148
## 15          WILDFIRE         75      911   986
```

Here, a new data frame "pubh_evt_a" is created, showing the fatalities, injuries and their sum as a function of the storm type.

From the visualization of this data set, I generate a new data set "d_c", which combines the information of "FATALITIES" and "INJURIES" in one column. To do this and do not mix the numbers, I generate a new column "Type of Cases".

With this form of the data set, it is much easier to generate the type of plot to be showed in the section Results.

The code for the new reordered data set and the plot is the following:

```r
#Prepare Data for geom_col:
d_a<-data.frame(pubh_evt_a[1:10,1],pubh_evt_a[1:10,2],"Fatalities")
names(d_a)<-c("Storm Event","Number of Cases","Type of Cases")
d_b<-data.frame(pubh_evt_a[1:10,1],pubh_evt_a[1:10,3],"Injuries")
names(d_b)<-c("Storm Event","Number of Cases","Type of Cases")
d_c<-rbind(d_a,d_b)
head(d_c)
```

```
##      Storm Event Number of Cases Type of Cases
## 1        TORNADO            5633    Fatalities
## 2 EXCESSIVE HEAT            1903    Fatalities
## 3      TSTM WIND             504    Fatalities
## 4          FLOOD             470    Fatalities
## 5      LIGHTNING             816    Fatalities
## 6           HEAT             937    Fatalities
```

```r
#Setting the colour-pallete
red<-rgb(.95,.5,.5,1)
```

```
lred<-rgb(.95,.25,.25,.75)
orange<-rgb(.95,.45,.25,.7)
lorange<-rgb(.95,.35,.15,.7)

#Creating the diagramm
gp1<-ggplot(data=d_c,aes(x=d_c[,1],y=d_c[,2],colour=d_c[,3],fill=d_c[,3]))+
geom_col()+
scale_x_discrete(name="Storm Event",limits=head(as.character(pubh_evt_a$EVTYPE),10)[10:1])+
scale_y_continuous(name="Number of Injuries and Fatalities")+
scale_colour_manual("", values = c(lred, lorange),
                    labels=c("Fatalities","Injuries"))+
scale_fill_manual("", values = c(red, orange),
                  labels=c("Fatalities","Injuries"))+
coord_flip()+
labs(caption="Data: NATIONAL WEATHER SERVICE INSTRUCTION 10-1605, AUGUST 17, 2007, NOAA.")+
theme(plot.caption = element_text(size=5))+
ggtitle("Impact of Storm-Event on Injuries and Fatalities\n Across the USA, 1950-2007")


gp2<-ggplot(data=d_c,aes(x=d_c[,1],y=d_c[,2],colour=d_c[,3],fill=d_c[,3]))+
geom_col(position="fill")+
scale_x_discrete(name="Storm Event",limits=head(as.character(pubh_evt_a$EVTYPE),10)[10:1])+
scale_y_continuous(name="Ratio of Injuries and Fatalities")+
scale_colour_manual("", values = c(lred, lorange) ,labels= c("Fatalities","Injuries"))+
scale_fill_manual("", values = c(red, orange),
                  labels=c("Fatalities","Injuries"))+
coord_flip()+
labs(caption="Data: NATIONAL WEATHER SERVICE INSTRUCTION 10-1605, AUGUST 17, 2007, NOAA.")+
theme(plot.caption = element_text(size=5))
#ggtitle("Impact of Storm-Event on Injuries and Fatalities\n Across the USA, 1950-2007")
```

2.2 Data Concearning Economic Damage and Event Type

Following section 2.1, the data concerning economic damage are analysed as follows:

```
econ_evt_a<-econ_evt %>%
  group_by(EVTYPE)%>%
  summarize(PROPDMG=sum(PROPDMG,na.rm=1),
            CROPDMG=sum(CROPDMG,na.rm=1),
            TOTAL=PROPDMG+CROPDMG) %>%
  arrange(desc(TOTAL))
head(econ_evt_a,15)
```

```
## # A tibble: 15 × 4
##              EVTYPE      PROPDMG     CROPDMG        TOTAL
##              <fctr>        <dbl>       <dbl>        <dbl>
## 1             FLOOD 144657709807  5661968450 150319678257
## 2 HURRICANE/TYPHOON  69305840000  2607872800  71913712800
## 3           TORNADO  56937160779   414953270  57352114049
## 4       STORM SURGE  43323536000        5000  43323541000
## 5              HAIL  15732267048  3025954473  18758221521
## 6       FLASH FLOOD  16140812067  1421317100  17562129167
## 7           DROUGHT   1046106000 13972566000  15018672000
```

```
## 8          HURRICANE  11868319010  2741910000  14610229010
## 9        RIVER FLOOD   5118945500  5029459000  10148404500
## 10         ICE STORM   3944927860  5022113500   8967041360
## 11    TROPICAL STORM   7703890550   678346000   8382236550
## 12      WINTER STORM   6688497251    26944000   6715441251
## 13         HIGH WIND   5270046295   638571300   5908617595
## 14          WILDFIRE   4765114000   295472800   5060586800
## 15         TSTM WIND   4484928495   554007350   5038935845
```

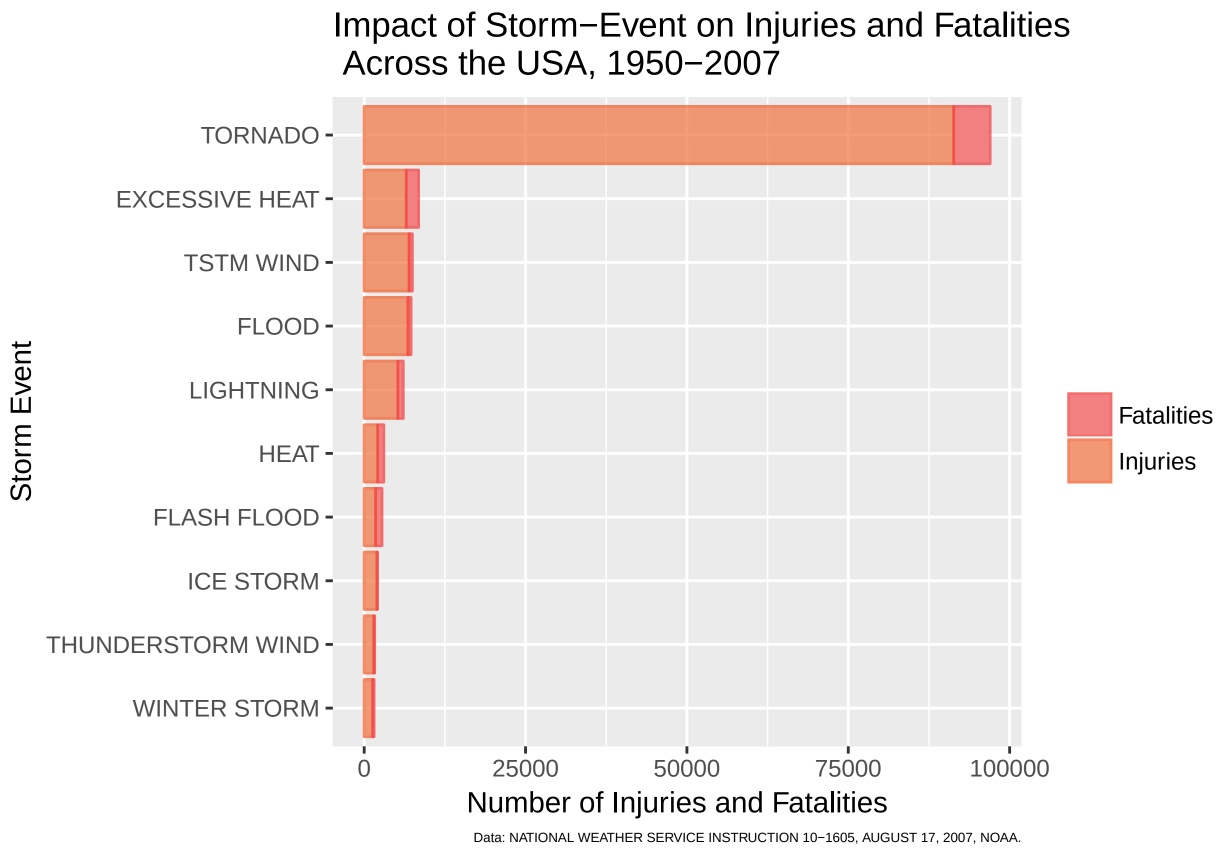
The visualization is also prepared as in section 2.1:

```r
#Prepare Data for geom_col:
d_a1<-data.frame(econ_evt_a[1:10,1],econ_evt_a[1:10,2],"Fatalities")
names(d_a1)<-c("Storm Event","Damage Amount","Type of Damage")
d_b1<-data.frame(econ_evt_a[1:10,1],econ_evt_a[1:10,3],"Injuries")
names(d_b1)<-c("Storm Event","Damage Amount","Type of Damage")
d_c1<-rbind(d_a1,d_b1)

green<-rgb(.5,.75,.5,1)
lgreen<-rgb(.5,.75,.25,.75)
yellow<-rgb(.75,.75,.25,.9)
lyellow<-rgb(.75,.75,.15,.9)

gp3<-ggplot(data=d_c1,aes(x=d_c1[,1],y=d_c1[,2],colour=d_c1[,3],fill=d_c1[,3]))+
geom_col()+
scale_x_discrete(name="Storm Event",limits=head(as.character(econ_evt_a$EVTYPE),10)[10:1])+
scale_y_continuous(name="Amount of Economic Damage in USD")+
scale_colour_manual("", values = c(lgreen, lyellow),
                    labels=c("Property Damage","Crop Damage"))+
scale_fill_manual("", values = c(green, yellow),
                  labels=c("Property Damage","Crop Damage"))+
coord_flip()+
labs(caption="Data: NATIONAL WEATHER SERVICE INSTRUCTION 10-1605,
     AUGUST 17, 2007, NOAA.")+
theme(plot.caption = element_text(size=5))+
ggtitle("Impact of Storm-Event on Economic Damage\n Across the USA, 1950-2007")


gp4<-ggplot(data=d_c1,
            aes(x=d_c1[,1],
                y=d_c1[,2],colour=d_c1[,3],
                fill=d_c1[,3]))+
geom_col(position="fill")+
scale_x_discrete(name="Storm Event",limits=head(as.character(econ_evt_a$EVTYPE),10)[10:1])+
scale_y_continuous(name="Ratio of Economic Damage in USD")+
scale_colour_manual("", values = c(lgreen, lyellow),
                    labels=c("Property Damage","Crop Damage"))+
scale_fill_manual("", values = c(green, yellow),
                  labels=c("Property Damage","Crop Damage"))+
coord_flip()+
labs(caption="Data: NATIONAL WEATHER SERVICE INSTRUCTION 10-1605,
     AUGUST 17, 2007, NOAA.")+
theme(plot.caption = element_text(size=5))
#ggtitle("Impact of Storm-Event on Economic Damage\n Across the USA, 1950-2007")
```

3. Results

3.1 Storms and Public Health - Tornadoes as the most harmful storm type.

```
gp1
```



The diagram shows the number of cases of fatalities and injuries caused by the different types of storm.

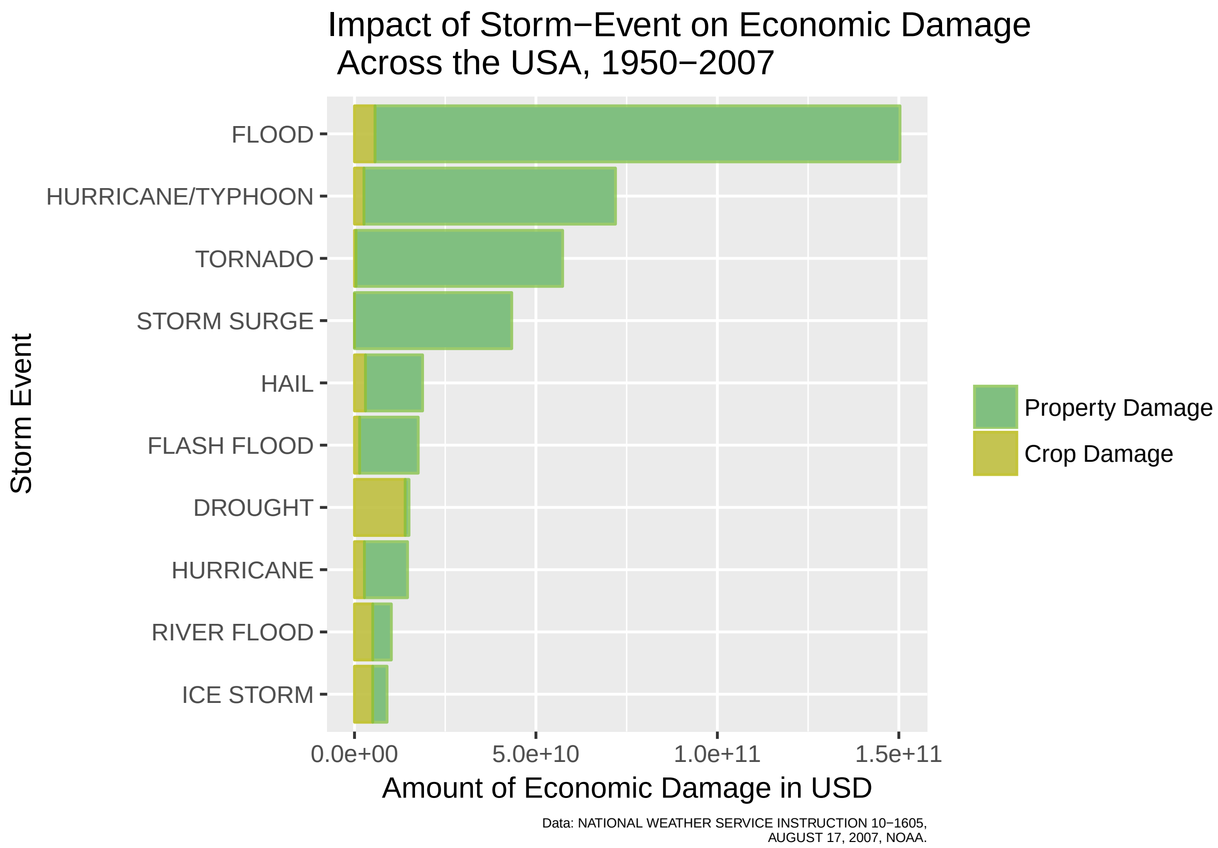
After these data, the tornadoes represent by far the most harmful type of storm, regarding both injuries and fatalities.

It is surprising, that the number of injuries from tornadoes exceeds all others by a factor of more than 15.

After the tornadoes, a group of disasters like excessive heat, thunderstorm winds, floods and lighting are responsible for at least 6.000 and 8.500 injuries and fatalities each from 1950 to 2007.

3.2 Storms and Their Economic Impact - Floods are responsible for the most of the damage.

```
gp3
```



As it can be seen in the presented diagram, floods are responsible for an estimate of nearly 0,15 trillion USD from 1950 to 2007.

Its damaging potential is specially threatening towards damages in property.

After the floods, a group of hurricanes, tornadoes and storm surges are responsible for economic losses between 400 to 700 billion USD each.

Again, the most affected value comes from damages in property.

It is remarkable, that among the most economically harmful disasters, the droughts are placed 7th., affecting above all crops. All other top ten damages come mostly from property damages.